Libro Metodológico del
Instituto Nacional de Estadística y Censos

Metodología

para transformar registros administrativos en registros estadísticos

Dirección
de Registros
Administrativos

Abril, 2022

INEC | Buenas cifras, **mejores vidas**

Instituto Nacional de Estadística y Censos - INEC

Autoridades del INEC:

Roberto Castillo
Director Ejecutivo
Jorge García
Subdirector General

Julio Muñoz
Coordinador General Técnica de Producción Estadística

Paúl Benavides
Director de Registros Administrativos

Autores: [1]
David Caín U.
Ángel Chiluisa V.
Verónica Figueroa V.
Carolina Guerrero T.
Sonia Herrera S.
Ruth Tenesaca S.
Marco Viteri R.

Citar como:
INEC (2022) Metodología para transformar registros administrativos en registros estadísticos. Instituto Nacional de Estadística y Censos (INEC), Quito, Ecuador.

Juan Larrea N15-36 y José Riofrío. Casilla postal 135 C / Telf: (02) 2544 - 326 / 2529 – 858

[1]Los autores agradecen los aportes y comentarios de los colegas: Víctor Espinoza y Yandre Jaime.

Contenido

Índice de ilustraciones

Índice de Tablas

Introducción

Los registros administrativos generados por las instituciones públicas y privadas, son una alternativa para ampliar y mejorar las estadísticas oficiales basadas en censos o encuestas (con levantamiento en campo), gracias a la captación continua, veraz y de bajo costo. Considerando que los registros administrativos son gestionados por entidades no estadísticas, cuando se trata de darles uso estadístico, deben pasar por un proceso de transformación previo a su análisis, de modo que las unidades de observación y las variables satisfagan necesidades de naturaleza estadística.

El uso de estos datos para fines estadísticos se remonta a finales de la década de los sesenta, cuando los institutos de estadística de los países nórdicos empezaron a utilizarlos en la generación de sus estadísticas, llegando a producir censos de población completos (United Nations Economic Commission for Europe, 2007).

En América Latina, los profesores Wallgren (2012) han impulsado el aprovechamiento estadístico de los registros administrativos, llegando a describir un sistema de registros, una metodología para el aprovechamiento de registros que forje las bases para crear el sistema, así como la definición de nueva terminología para una teoría que goce de aceptación general. Es así que México ha desarrollado un proceso estándar para el aprovechamiento de registros administrativos, estableciendo 6 fases para la generación de estadísticas utilizando registros administrativos (INEGI, 2012).

En este marco, el Instituto Nacional de Estadística y Censos (INEC) desde el año 2014 con la creación de la Dirección de Registros Administrativos (DIRAD), ha trabajado en el aprovechamiento estadístico de los registros administrativos, teniendo como objetivo central, recopilar, procesar e integrar datos provenientes de distintas fuentes administrativas para posteriormente utilizarlos en la producción estadística.

Este documento establece las macro actividades referenciales a ejecutarse en el proceso de transformación de registros administrativos en estadísticos, que posteriormente sirvan para la generación de operaciones estadísticas basadas en registros administrativos. Se divide en 3 secciones: La primera contiene el Marco conceptual y referencial. La segunda sección presenta el proceso de Transformación de registro administrativo en estadístico. Finalmente, la tercera sección describe las principales conclusiones y recomendaciones.

1. Marco conceptual y referencial

1.1. Registros administrativos y su potencial estadístico

Por la naturaleza con la que son levantados los registros administrativos, estos presentan las siguientes características (DANE, 2009):

- Captación continúa de información asociada a actividades específicas sobre personas o entidades.
- Contienen información a nivel individual (personas o entidades), y abarcan segmentos considerables de la población.
- Son creados para fines administrativos, y es posible utilizarlos como fuente de información estadística.

Así también, con base en la experiencia del INEC, se ha evidenciado que un registro administrativo para fines estadísticos se encuentra en formato digital y estructurado (tabla de datos), y es recopilado por medio de escenarios seguros de transferencia de información, mismos que son preestablecidos entre la institución fuente (proveedor) y el INEC.

Una vez que los registros administrativos se han transformado en estadísticos, estos pueden tomar los siguientes usos:

- **Mejorar la precisión:** Validar y mejorar la identidad de los entes, permitiendo mejorar el nivel de integración con nuevas fuentes de información.
- **Mejorar la calidad:** Corregir e imputar los datos inconsistentes de las operaciones estadísticas, siempre que los datos de registros administrativos tengan buena calidad o provengan de una fuente primaria.
- **Mejorar la completitud y cobertura:** Completar valores faltantes en las variables, y mejorar o ampliar la cobertura de las operaciones estadísticas.
- **Generar nuevos estudios:** Añadir nuevas variables a operaciones estadísticas, y generar nuevos análisis con datos exclusivamente de registros administrativos.

Por otra parte, los registros administrativos tienen sus ventajas y desventajas cuando se tratan de utilizarlos en la producción estadística; entre sus ventajas están el contar con una cobertura completa de la población de registro y una adecuada cobertura geográfica que permite obtener información para áreas geográficas pequeñas; el disponer de los datos a nivel de microdato (datos individuales) y, con mayor oportunidad dado que se levanta la información cuando ocurre el hecho o se inicia una actividad (captación continua y oportuna); además, permite la reducción de

costos en el levantamiento ya que la información la levanta la institución fuente, así como reduce los errores de precisión y no respuesta, porque los datos se generan cuando un usuario accede/requiere un bien o servicio (calidad); así también, permiten generar paneles longitudinales para estudios longitudinales, e integrar nuevas fuentes de información para estudios transversales (Wallgren & Wallgren, 2012).

Entre las desventajas se pueden citar la baja calidad en variables de menor importancia administrativa, dado que no tienen un adecuado control al momento de levantar los datos (no tienen mucho interés administrativo), esto tomando en cuenta que la información es generada con base en las prioridades y políticas de la institución fuente, las cuales no se ajustan a las necesidades de la producción estadística, además, existe aún la ausencia o desactualización de registrar un hecho por falta de incentivo/obligatoriedad, lo que afecta a la cobertura y oportunidad de la información y, requieren de tareas de limpieza (pre-procesamiento) de las bases de datos, lo cual demanda tiempo y recursos tanto humanos como tecnológicos (Wallgren & Wallgren, 2012).

Finalmente, se debe considerar que debido a que los registros administrativos son levantadas por diversas fuentes (instituciones) y no necesariamente para fines estadísticos, es necesario evaluar su estado actual (materia prima) previo a ejecutar la transformación en registros estadísticos; la metodología que dispone el INEC para evaluar a un registro administrativo lo hace a tres componentes, mismos que tienen sus respectivos indicadores y métricas, estos son: metadato descriptivo o referencial del registro administrativo, metadato estructural de las variables del registro y, microdato (INEC, 2019).

1.2. Referencias del procesamiento estadístico de registros administrativos

Dado que los registros administrativos responden a leyes y reglamentos que deben cumplir las instituciones para su operación, no es recomendable generar directamente estadísticas a partir de estos, sino que deben transformarse para pasar de registros administrativos a estadísticos (Ver Ilustración 1); el esquema propuesto por Wallgren & Wallgren (2007) tiene como entrada un dato administrativo, el cual debe pasar por un procesamiento para obtener como producto final, un dato estadístico.

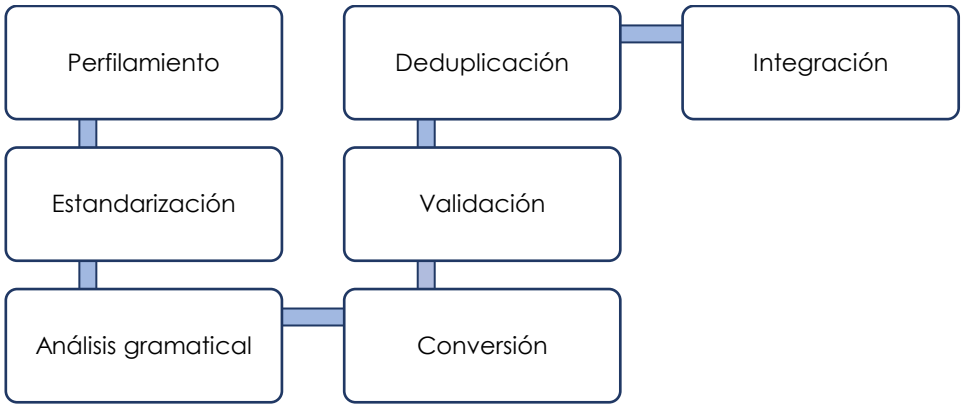
Ilustración 1: Transformación de registros administrativos a estadísticos - Wallgren

Fuente: Wallgren & Wallgren (2007)

Así mismo, el procesamiento de datos administrativos tiene como procesos transversales la confidencialidad y el aseguramiento de la calidad (Wallgren & Wallgren, 2012).

Para el caso europeo se ha revisado el proceso de transformación de datos administrativos en estadísticos de la Oficina Estadística de la Unión Europea (2012), mismo que establece siete etapas, que se enlistan en la Ilustración 2.

Ilustración 2: Transformación de registros administrativos a estadísticos – Eurostat



Fuente: Economic Commission for Europe, 2012

Así también, el Instituto Nacional de Estadística y Geografía (INEGI) de México, en el documento "Proceso estándar para el aprovechamiento de registros administrativos" (INEGI, 2012), establece en 6 fases consecutivas y 2 transversales el proceso para aprovechar estadísticamente los registros administrativos (Ver Ilustración 3).

Ilustración 3: Fases para la generación de estadísticas por aprovechamiento de registros

Fuente: INEGI, 2012

Las fases 4 a la 6 se repiten en forma cíclica de acuerdo con lo establecido en las fases 1 y 2, mientras que las fases de control administrativo y la gestión de la calidad, son transversales para todas las fases.

Es importante citar al Modelo de Producción Estadística (MPE) del Ecuador (INEC, 2016), que es un modelo estándar para la generación de operaciones estadísticas del Sistema Estadístico Nacional (SEN), incluidas las estadísticas basadas en registros. Este modelo está conformado por tres niveles: fases, etapas, y actividades, con ocho fases y dos macro procesos transversales. En la Ilustración 4 se presenta el MPE a nivel 2.

Ilustración 4: Modelo de Producción Estadística - Ecuador

Fuente: INEC (2016)

2. Transformación de registro administrativo en estadístico

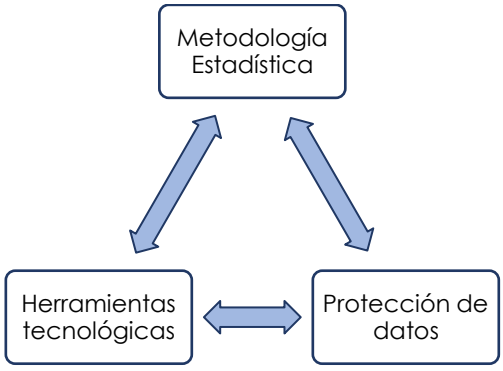
La transformación de un registro administrativo se ejecuta posterior a las fases de Planificación, Diseño y Construcción (Ciclo de preparación), propias del MPE; por lo que en las primeras fases y bajo un amparo legal, se deben definir los siguientes temas principales: listado de tablas, variables y metadatos a transferir, periodo (fechas), y medio tecnológico seguro de trasferencia.

Es importante señalar que las instituciones fuente captan datos en sus registros administrativos a través de distintos medios (papel, sistemas, etc.), que para su uso en la producción estadística deben almacenarlos en "tablas de datos", procesarlos y finalmente utilizarlos en la producción estadística. Por lo tanto, se tendrán las siguientes etapas: Registros Administrativos, Registros Estadísticos y Operaciones Estadísticas basadas en registros administrativos.

Los siguientes tres elementos interactúan entre sí (Ver Ilustración 5), y son esenciales para el proceso de transformación de un registro administrativo en estadístico, por lo que es necesario conformar un equipo multidisciplinario donde participen metodólogos e informáticos, y así recopilar y procesar adecuadamente los registros que serán utilizados en la generación de Operaciones Estadísticas.

- **Metodología estadística. -** Teoría estadística y criterios metodológicos para transformar registros administrativos en estadísticos, donde se establece una secuencia de procedimientos para el manejo de datos.
- **Herramientas tecnológicas. -** Hardware y software para recopilación y procesamiento seguro y periódico de grandes volúmenes de datos.
- **Protección de datos. -** Directrices y herramientas con las cuales se cuidará el sigilo de los datos en la entrada, el procesamiento y la salida; dado que se cuentan con datos de identificación (confidenciales) individual.

Ilustración 5: Elementos básicos para el aprovechamiento estadístico de registros



Elaboración: INEC

Como se mencionó, el documento se centra en las fases de Recopilación y de Procesamiento,[2] con las cuales comienza todo el ciclo de operación, teniendo a la protección de los datos como fase transversal (Verl Ilustración 6); las demás fases se deben considerar del MPE vigente en Ecuador.

Ilustración 6: Modelo de Producción Estadística con Registros Administrativos - Ecuador

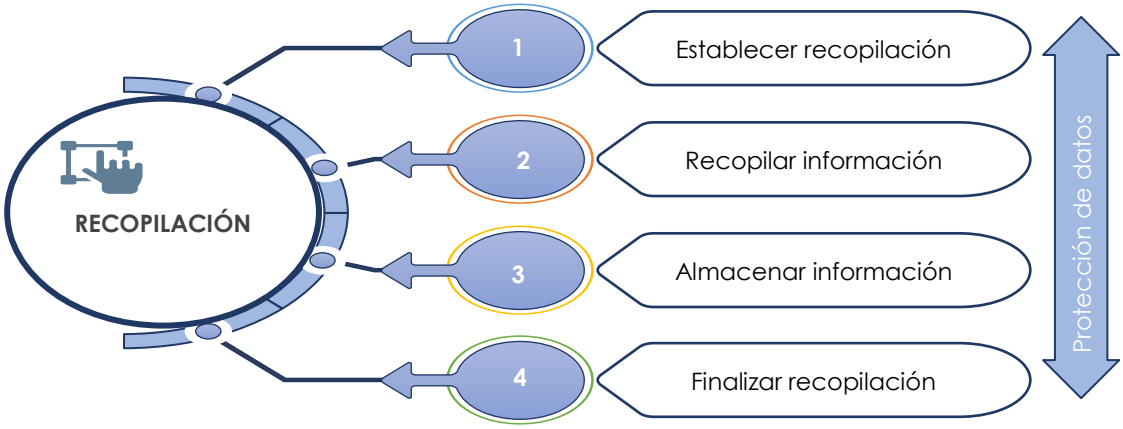
Elaboración: INEC

2.1. Recopilación

Los datos que son captados por la institución/área de acuerdo al ámbito de su competencia, son transferidos a las instituciones/áreas respectivas para su procesamiento con fines estadísticos. Esta fase contiene cuatro macro actividades que se describen a continuación en la Ilustración 7.

[2] La fase de recolección toma el nombre de recopilación debido a que en esta fase se recopilan bases de datos desde distintas fuentes, y no se recolectan datos en campo como en las encuestas o censos tradicionales.

Ilustración 7: Macro actividades para la recopilación de información



Elaboración: INEC

2.1.1 Establecer la recopilación (Captación de datos)

Cuando son datos de fuentes externas se considerarán las condiciones legales acordadas entre las partes y, si son levantadas por la propia Institución, se debe establecer el contexto técnico para la captura y almacenamiento de los mismos, conforme a la fase de Planificación del MPE.

En esta macro actividad se establecen los detalles técnicos para la transferencia de las bases de datos desde la institución fuente hacia la requirente, para lo cual se considerarán lo establecido en el acuerdo legal firmado entre las partes. Las principales actividades a ejecutar están el confirmar fechas (calendario) de transferencia, de acuerdo a la disponibilidad de datos en la fuente administrativa y establecer/confirmar y probar el escenario tecnológico de transferencia, velando por la protección de los datos, de acuerdo a lo que se haya establecido en las fases previas del MPE.

Los escenarios de transferencia se conforman por la combinación de medios y modos de transferencia:

Medios de transferencia:

- **Internet:** Red informática mundial, descentralizada, formada por la conexión directa entre computadoras mediante un protocolo especial de comunicación (RAE, 2021).
- **Red de datos:** Infraestructuras creadas para transmitir información a través del intercambio de datos, mismas que podrían ser por:
    - **DbLink:** Objeto en Oracle (base de datos) para realizar conexión de una base de datos a otra; su principal objetivo es ocultar el detalle de la

conexión, facilitando acceso a los recursos disponibles en la base de datos, sea propia o externa (Oracle, CREATE DATABASE LINK, 2022).

- **SFTP:** Protocolo de red para la transferencia segura de archivos, basado en la arquitectura cliente-servidor (Bravo, 2015), donde un equipo cliente se puede conectar a un servidor para descargar archivos desde él o para enviar archivos.
- **Red de datos gubernamentales:** Infraestructura o red de comunicación diseñada en Ecuador para la transmisión de información entre instituciones del gobierno.

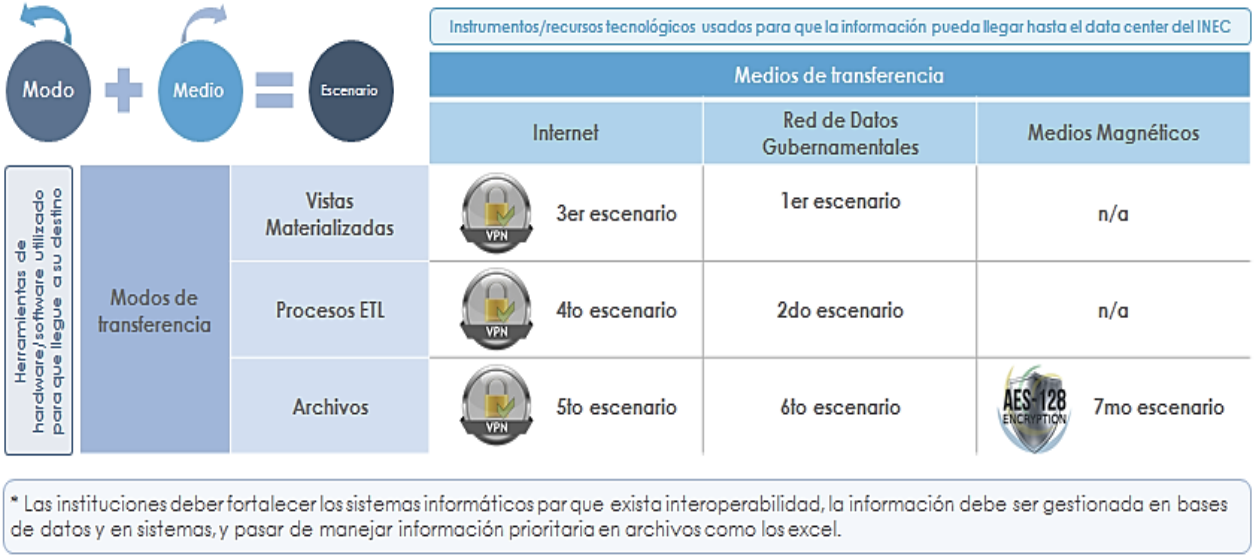
- **Medios magnéticos:** Dispositivo que utiliza materiales magnéticos para archivar información digital, tales como discos duros o CD que almacenan grandes volúmenes de datos en un espacio físico pequeño.

Modos de transferencia:

- **Vistas materializadas:** Se define como una vista común, que almacena el resultado de la consulta, es decir, la materializa, como un objeto persistente en la base de datos (Oracle, CREATE MATERIALIZED VIEW, 2022).
- **Procesos ETL:** Comprende tres fases: extraer (E. Extracción), transformar (T. Transformación) y cargar (L. Carga).
- **Archivos (fichero):** Unidad de datos o información almacenada en algún medio, que puede ser utilizada por aplicaciones de la computadora; su estructura depende del software en el que se genera el archivo de información, y por lo tanto el formato. Los ficheros pueden ser de tipo:
  - **Archivo plano:** Contiene texto en diferentes formatos (.sav, .txt, .csv, .xls, .xlsx).
  - **Backup:** Puede tener varios formatos que dependerán del motor de base de datos donde fue generado.

La Ilustración 8 presenta los escenarios de transferencia en el orden de prioridad deseado (con base en la experiencia del INEC) según las bondades que presentan en tiempo y recursos.

Ilustración 8: Escenarios de transferencia de información



Elaboración: INEC

- **1er escenario (vistas materializadas + red de datos gubernamental).** Se refiere a la creación de vistas materializadas en donde los datos son transmitidos por la red de datos gubernamental, estas reciben una carga de información inicial y al ser configuradas para su actualización automática, reciben únicamente datos nuevos o actualizados, reduciendo de esta manera el tamaño de los registros, tiempo de recepción y trabajo manual.
- **2do escenario (procesos ETL + red de datos gubernamental).** Se diseñan procesos ETL para la captación de datos, los cuales son transmitidos por la red de datos gubernamental. Al ser un proceso ETL no importa el origen de los datos como SGBD, archivos, etc.; que pueden ser transformados de manera automatizada al formato que el receptor determine.
- **3er escenario (vista materializada + internet + VPN).** Similar al 1er escenario, con la diferencia de que la transmisión de datos es a través de una VPN, que en muchos casos se utiliza una conexión por medio de un DbLink, con la que se accederá a la vista materializada.
- **4to escenario (proceso ETL + internet + VPN).** Similar al 2do escenario, con la diferencia en la construcción de la VPN, que permita acceder a los datos de origen que serán transformados mediante el ETL.
- **5to escenario (archivos + internet + VPN).** Transferencia de archivos donde el formato es indispensable para la recepción de la información, ya que depende de la codificación en la que se encuentren, y cómo fueron generados, es decir,

desde cuál herramienta fueron creados, siendo necesario que se disponga de la misma herramienta para su restauración.

- **6to escenario (archivos + red de datos gubernamental).** La transferencia de archivos es mediante el uso de la red gubernamental, y requiere de tareas manuales para su posterior restauración, lo cual dificulta la automatización de los procesos de recopilación.
- **7mo escenario (archivos + medios magnéticos + encriptación).** Este escenario ya no es tan común, ya que implica entrega de datos por medios magnéticos, y la total dependencia de los algoritmos de encriptación o des-encriptación.

2.1.2 Recopilar la información

Habiéndose confirmado y probado el escenario de transferencia entre el proveedor y el requirente, se ejecuta la transferencia, en la cual las contrapartes de cada lado deberán notificar las características de la información que se transfieren. Las principales características a verificarse son: nombre de la tabla, tamaño del archivo, número y listado de variables de la tabla (incluye descripción), catálogos de las variables categóricas, número de registros que contiene la tabla y, fecha de corte de los datos.

En caso de no contar con estos detalles, se los debe levantar, ya que es el punto de partida para mejorar los procesos internos o retroalimentar al proveedor de información.

2.1.3 Almacenar la información

Los datos recibidos deben ser almacenados y ejecutados una verificación inicial, donde se revisará que cumpla las condiciones establecidas en el instrumento legal; en caso de existir novedades, no se almacenará, sino que deberá solicitar una retroalimentación al proveedor, para posteriormente almacenar los datos debidamente corroborados.

Para este proceso se requiere de herramientas tecnológicas que faciliten la recopilación, almacenamiento y perfilamiento de los datos. Así también, el custodio de la información en la institución/área requirente, una vez inspeccionado lo recibido (literal b), debe almacenar la BDD en un repositorio destinado para tal fin.

2.1.4 Finalizar la recopilación

Ya con los datos recopilados y almacenadas en la institución/área requirente en los ambientes de captación, se debe registrar o actualizar periódicamente el registro con

el "Histórico de Recopilación de Registros Administrativos"[3], con la siguiente información: fecha de corte y transferencia, nombre de la institución fuente (proveedora), nombre y descripción de la base de datos (población registrada, unidad registrada, etc.), tamaño (volumen) de la base de datos, cantidad (número) de registros, entorno de almacenamiento (lugar de almacenamiento o restauración), escenario de transferencia de datos (medio y modo) y, código hash (para comprobar la integridad de la BDD, solo en escenario 7).

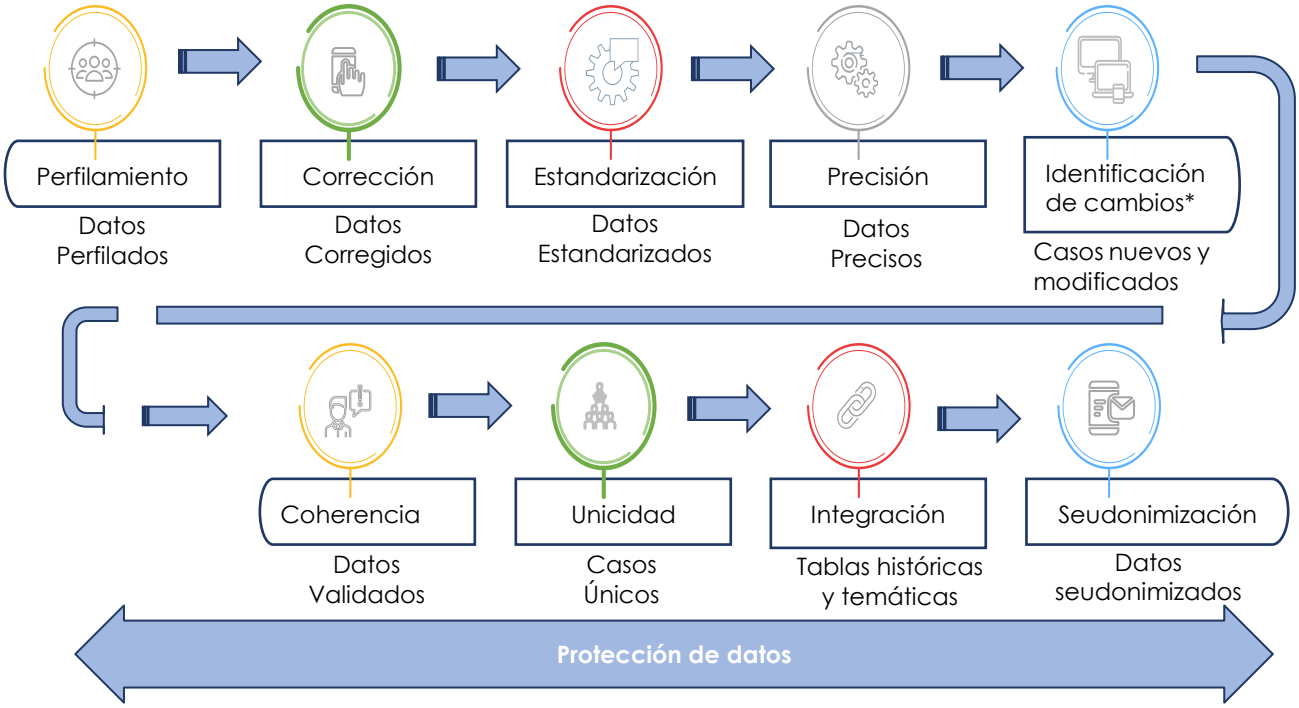
Finalmente, se debe realizar un pre-perfilamiento para identificar y eliminar los caracteres especiales que provoquen saltos de línea en los registros, para posteriormente cargar en el entorno de procesamiento.

2.2 Procesamiento

Una vez recopilados los datos y sus respectivos metadatos, se procede con la transformación a registro estadístico, para lo cual se han establecido nueve macro actividades referenciales (Ver Ilustración 9), que se aplicarán principalmente a datos provenientes de fuentes externas y que contengan información sensible y/o confidencial. Es importante que en cada macro actividad se generen reportes que permitan monitorear las acciones ejecutadas y las novedades en los datos.[4]

[3] Tabla tipo inventario que almacena los metadatos de los registros administrativos en cada entrega (transferencia).
[4] A partir de los resultados obtenidos en las distintas macro actividades se puede realizar la Evaluación del estado actual del registro administrativo a nivel de microdato (INEC, 2019); con estos resultados se podrá cuantificar su calidad y determinar los aspectos a mejorar/fortalecer.

Ilustración 9: Macro actividades del procesamiento de información



** La (5) identificación de cambios está relacionada con la (8) Integración (construcción de tablas históricas).*

Elaboración: INEC

2.2.1 Perfilamiento

Aquí se identifican errores o novedades en la estructura de los datos, por cada variable, teniendo como objetivo, corroborar que los datos recibidos se ajusten a los formatos establecidos.

Para esta macro actividad se deben agrupar a las variables según el tipo de dato (Ver Ilustración 10).[5]

[5] Clasificación que toma de referencia a los profesores Wallgren & Wallgren (2012).

Ilustración 10: Clasificación de variables para procesamiento

Elaboración: INEC

En función de esta agrupación, se deben generar sintaxis (scripts/algoritmos) de ejecución o reporte, que permitan identificar las novedades por cada tipo de variable. Los principales scripts para evaluar los datos son de tipo:[6] cédula, RUC, fecha, número, número decimal, cadena de texto, correo electrónico, teléfono, etc. Cada script o reporte debe incluir un catálogo de errores que permita identificar y registrar los distintos tipos de novedades. Por ejemplo, para datos tipo cédula, que se puedan clasificar los errores/novedades como válidos, vacío, incorrectos, etc. (Ver las dos primeras columnas de la Tabla 1). Como resultado de esta macro actividad salen datos perfilados (identificadas las novedades) y que serán utilizados en la corrección.

2.2.2 Corrección

Aquí se corrigen los errores "de forma" identificados en el perfilamiento, mismos que dependen del tipo de dato y tipo de error catalogados en el punto anterior (los criterios de corrección dependen de cada tipo de dato). Así, un dato puede tener más de una novedad registrada, por ejemplo, una cédula que tiene espacios en blanco (catalogo error = CC07) y tiene 9 dígitos en lugar de 10 (catálogo error = CC04). De esta macro actividad salen datos corregidos (corregidas las novedades de forma identificadas en el perfilamiento).

La Tabla 1 muestra un ejemplo de reporte de perfilamiento (2 primeras columnas) y sus respectivas correcciones para los datos tipo "cedula".

[6] Cada script debe tener reglas específicas según su tipo de dato, por ejemplo, para la cédula: sólo números, longitud de 10 campos, aplicar el logaritmo de base 10, etc.

Tabla 1: Correcciones aplicadas a los datos tipo "cedula"

DESCRIPCIÓN DEL ERROR	CÓDIGO ERROR	CORRECCIÓN DE ERROR ENCONTRADO
Cédula correcta: 10 dígitos y pasó el dígito verificador (módulo 10)	CC00*	Retorna el mismo valor (cédula).
Cédula con valor en nulo o vacíos	CC01	Retorna valor vacío.
Cédula con más de 10 dígitos	CC02	Retorna el mismo valor de entrada (no corrige).
Cédula con menos de 9 dígitos	CC04	Añade el 0 al inicio, en caso de pasar el algoritmo "módulo 10" corrige con el 0 al inicio, caso contrario retorna la cédula con los mismos dígitos de entrada.
Contiene caracteres especiales y/o letras en número	CC06	Ya que la cédula acepta únicamente números, elimina caracteres especiales y letras, y reemplaza O por 0 y S por 5.
Contiene espacios en blanco en el intermedio	CC07	Elimina los espacios en blanco de los intermedios.

*/ Todo código que contiene 00 indica que tiene un valor correcto.
Elaboración: INEC

Los códigos CC* son definidos para uso interno del INEC, se pueden utilizar los mismos o asignar de acuerdo a las preferencias y necesidades institucionales.

2.2.3 Estandarización

Se homologan las distintas categorías de las variables cualitativas que vienen desde la fuente a las del estándar nacional o internacional. Es importante contar con la descripción o concepto de cada categoría, con los cuales se atarán (matriz de correspondencia) a los del estándar. Las categorías estándar toman el nombre de catálogo "padre", al cual se atan/homologan los catálogos "hijos" (cada padre puede tener n hijos, pero cada hijo tendrá un sólo padre).

El INEC cuenta con una herramienta tecnológica llamada METADEC con la cual gestiona los metadatos de los registros administrativos, misma que contiene los metadatos estructurales de las variables, y está basada en el estándar SDMX (Statistical Data and Metadata Exchange) [7]. En esta herramienta se crea y configura lo siguiente:

[7] Los metadatos estructurales se utilizan para identificar, describir formalmente y entender a los microdatos; entre sus principales elementos constan: concepto, tipo de dato, formato o clasificaciones.

- **Agencias:** Nombre de la institución fuente/proveedor y la agencia estándar/padre (INEC).
- **Estructuras de datos:** Agrupa a las variables (conceptos) que serán utilizadas para estandarizar una variable específica.
- **Variables de registros administrativos:** Listado de variables de cada registro administrativo.
- **Conceptos:** Es lo que hace a una variable descifrable o entendible.

  Catálogos / categorías: Permiten almacenar la información codificada para luego ser asociada a cada concepto (de la variable).

  La Tabla 2 presenta un ejemplo de la matriz de correspondencias con los catálogos / categorías para estandarizar la variable sexo.

Tabla 2: Estandarización de la variable sexo – METADEC

CATEGORÍAS	DESCRIPCIÓN	ESTÁNDAR	CLAVE SUBROGADA	PERIODO
M	Hombre	1	1	2019-01-01 – 2999-12-31
F	Mujer	2	2	2019-01-01 – 2999-12-31
N	No registra	99	3	2020-12-31 – 2999-12-31

En periodo se deja como 2999-12-31 para indicar que está activo, caso contrario se registra una fecha hasta la cual estuvo vigente ese catálogo.

Elaboración: INEC

Los catálogos hijos (categorías) se vinculan con sus correspondientes catálogos padres (estándar) a través de la descripción; para estos efectos, la clave subrogada es el código único asignado a los conceptos/raíz de cada categoría, mismo que controla las variaciones y vigencia de una categoría en el tiempo (Ver Tabla 3). Cada categoría tendrá su periodo o vigencia (fecha inicio y fecha fin), identificándose a las que están activas con fecha fin igual a 2999-12-31 (periodo abierto).

Tabla 3: Uso de clave subrogada en la estandarización

CATEGORÍAS	ESTÁNDAR	CLAVE SUBROGADA	PERIODO
Quito	1701	1	2008-01-01 – 2999-12-31
UIO	1701	1	2008-01-01 – 2999-12-31
Santo Domingo	1706	2	2008-01-01 – 2009-06-30
Santo Domingo	2301	2	2009-06-30 – 2999-12-31

Elaboración: INEC

Como resultado de esta macro actividad salen datos estandarizados.

2.2.4 Precisión

Se verifica que la identidad de un ente (persona o empresa) corresponda efectivamente a la real (paso 1, validación y corrección), y en caso de no serlo, se

procederá a recuperar (paso 2). Para esto, es necesario contar con datos de una fuente primaria (ejemplo: para personas, los datos de cedulados del Registro Civil),[8] con la cual se compararán los datos de identificación provenientes del registro administrativo que se esté procesando.

Para la validación (paso 1) y la recuperación (paso 2) de la identidad, se deben considerar métodos de emparejamiento determinísticos y probabilístico. El determinístico se realiza cuando las variables de identificación empatan al 100% entre dos fuentes, en tanto que el probabilístico cuando el emparejamiento tiene cierto grado de pertinencia entre dos fuentes, por lo que se deben considerar umbrales de similitud que permitan con cierto nivel de pertinencia, determinar los casos como "verdaderos"; esto debido a que los datos no empatan al 100% de similitud.[9]

Dentro de esta macro actividad, se deben generar dos variables técnicas:

- Una que registre el nivel de certeza o match de la identidad (valida_id), dependiendo del porcentaje de similitud de las cadenas de textos, como producto de las distintas combinaciones de las variables auxiliares (nombres, apellidos, fecha nacimiento, lugar de nacimiento y sexo) y,
- Otra que resuma los resultados del punto anterior en 3 grupos (verifica_id): 1) válidos verdaderos, aquellos casos cuya información es consistente con la fuente primaria (identidad confirmada), 2) válidos falsos, aquellos casos en donde la variable de identificación es igual a la de la fuente primaria, sin embargo, el resto de variables no coinciden (identidad no confirmada) y, 3) "Null e incorrectos.

La ejecución de esta macro actividad se divide en dos pasos, el paso 1 toma como variable llave la clave primaria (cédula o RUC.) y el paso 2 toma como llave a las variables auxiliares (ejemplo para personas: nombres, apellidos, fecha nacimiento, lugar de nacimiento y sexo):

[8] Fuente primaria: Registro administrativo proveniente de la institución que por mandato legal genera el dato oficial.
[9] Por la experiencia con los datos en Ecuador, se han implementado la combinación de dos algoritmos de similitud: Jaro Winkler y N-grams, así también para determinar los umbrales se utiliza la teoría de match difuso (lógica difusa).

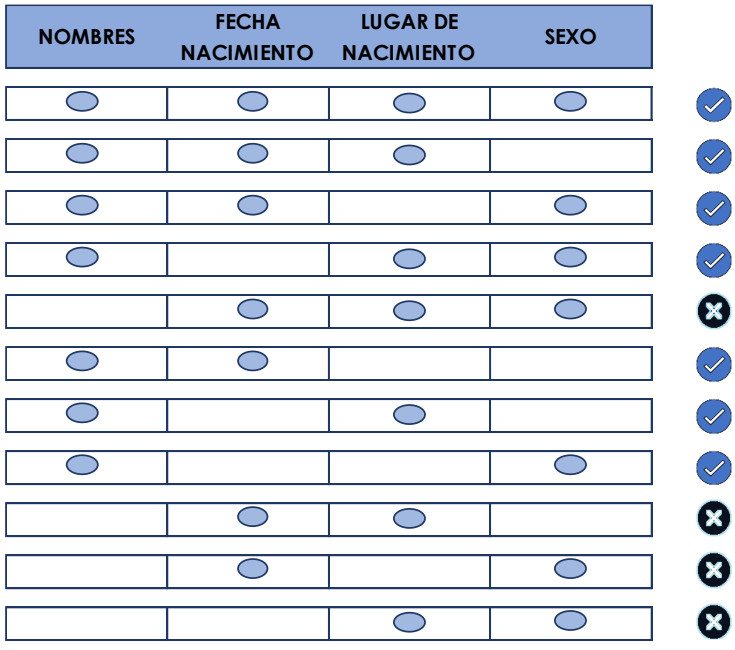
1. VERIFICACIÓN DE LA IDENTIDAD - ID CORRECTOS

Se inicia verificando la identidad de aquellos casos cuyo identificador único (cédula, RUC, etc.) pasaron como válidos en el macro proceso de Perfilamiento, para lo cual se siguen los siguiendo estos pasos:

a) Determinar las variables auxiliares que confirmen la identidad del ente, entre ellas: nombres y apellidos, fecha de nacimiento, lugar de nacimiento y sexo (para empresas u otras unidades de análisis, se debe buscar un símil).
b) Establecer las combinaciones con las cuales se confirmará la identidad del ente, apoyados de la teoría de las combinaciones $nCr = \frac{n!}{(n-r)!*r!}$, se calcula el total de combinaciones a ejecutar; se considerarán las combinaciones que estrictamente incluyan la variable Nombres o Razón Social, según sea el caso.

La Ilustración 11 muestra la aplicación de la teoría de combinaciones con cuatro variables, con las cuales se obtienen 1 grupo de 4 ($_4C_4$=1), 4 grupos de 3 ($_4C_3$=4) y 6 grupos de 2 ($_4C_3$=4) variables; así también, como se dijo anteriormente, de las once combinaciones calculadas, se aplicarán siete, dado que son las combinaciones que tienen a la variable Nombres.

Ilustración 11: Combinaciones con 4 variables (auxiliares) – para precisión



Elaboración: INEC

c) Se empareja con la fuente primaria utilizando el número de identificación único (clave primaria: cédula, RUC, etc.) y se corrobora la identidad con sus respectivas variables equivalentes.

   Como resultado se obtienen los casos considerados como 1 "válidos verdaderos", y 2 "válidos falsos".

2. RECUPERACIÓN DE LA IDENTIDAD: VÁLIDOS FALSOS, IDENTIFICACIÓN INCORRECTOS (INCLUYE NULOS)

Seguidamente, aquellos casos con identidad no confirmada del paso anterior (válidos falsos) más los nulos o no válidos, se proceden a la recuperación de la identidad, mediante emparejamiento con las variables de identificación auxiliar como: nombres y apellidos, fecha de nacimiento, lugar de nacimiento y sexo.

Para lo cual se ejecuta un emparejamiento con las variables de identificación auxiliar equivalentes de la fuente primaria, y se procede a recuperar el número de identificación único (clave primaria: cédula, RUC, etc.); similar a la validación, se utilizan también distintas combinaciones de las variables auxiliares.

Los casos que no fueron posible la recuperación por medio de algoritmos, se procede a recuperar manualmente utilizando herramientas oficiales que permitan consultar y validar la identidad.

Es importante registrar por cada procesamiento, el número de casos que almacenan las variables técnicas (verifica_id y valida_id), con los cuales se podrá dar seguimiento en el tiempo a la calidad de sus datos; así también, se debe considerar que a los registros administrativos considerados como "fuente primaria" no se ejecuta este proceso de precisión. Como resultado de esta macro actividad salen datos verificados frente a los datos de la fuente primaria.

2.2.5 Identificación de cambios

Considerando el crecimiento que pueda tener una tabla de datos en su volumen, por las modificaciones en el tiempo que sufra el dato de un ente, y con el objetivo de optimizar los recursos tecnológicos, se deben identificar los casos nuevos, modificados y, eliminados entre el tiempo t_n con relación a un tiempo anterior t_{n-1}, para que sólo aquellos casos nuevos y modificados continúen las siguientes fases del procesamiento, es decir, dejando a un lado los procesados en un periodo anterior (t_{n-1}).

Esta macro actividad se aplica cuando la fuente transfiere toda la Base de Datos Histórica, es decir, sin considerar únicamente aquellos que corresponden al último periodo.

Tabla 4: Casos que pueden encontrarse en una tabla en el periodo t_n y t_{n-1}

CASO	DESCRIPCIÓN
0=Igual	Casos que cruzan entre el periodo t_n y t_{n-1}, y tienen los mismos valores en las variables (número de control).
1= Nuevos	Casos que no cruzan entre el periodo t_n y t_{n-1}.
2= Modificados	Casos que cruzan entre el periodo t_n y t_{n-1}, pero tienen diferencias en los datos (número de control).
3= Eliminados	Casos que no cruzan entre el periodo t_n y t_{n-1}, es decir llegaron en t_{n-1} pero no en t_n.

Elaboración: INEC

Aquí deben compararse los datos del periodo actual (t_n) con la del periodo anterior (t_{n-1}), siendo necesario crear un código único llamado "número de control" con las variables que previamente se definan y, posteriormente haciendo uso de las variables llaves (cédula/RUC más las que hagan único al registro) se deben comparar los valores del número de control entre los dos periodos, para conocer si los datos que se emparejaron son los mismos o hubo algún cambio; así también los casos que no cumplieron con la condición serán considerados como: nuevos o eliminados, según sea el caso.[10] Así, en esta macro actividad se identifican los casos a seguir procesando (nuevos y modificados).

2.2.6 Coherencia

Aquí se realizan validaciones de datos entre variables de un mismo registro administrativo, permitiendo verificar que los datos de una variable con otra no tengan contradicciones, siendo necesario establecer variables a validar y variables validadoras, con los cuales se genera una matriz con las reglas de validación y su respectiva corrección (ver Tabla 5). De ahí que de esta macro actividad salen datos validados y del ser el caso corregidos.

[10] El "número de control" es un código (número) que se genera mediante operaciones matemáticas entre el código ASCII y la posición de cada letra.

Tabla 5: Validación y corrección de datos - ejemplo

ACCIÓN	CASO	VARIABLE VALIDADA	VARIABLE VALIDADORA	DESCRIPCIÓN
Validación	Caso 1	condicion_cedulado	fecha_fallecimiento	Los casos con fecha_fallecimiento diferente de null (que tengan una fecha válida) deben tener en condicion_cedulado 7 o 20 (fallecido).
	Caso 2	estado_civil	Edad	Una vez que la edad sea precisa, aquellos casos con edad <= 18 deben tener estado_civil "soltero"
Corrección	Caso 1	condicion_cedulado		Los casos que no cumplan con la condición establecida, deben cambiarse a 88 en condicion_cedulado, con la descripción de que están fallecidos.
	Caso 2	estado_civil		Los casos que no cumplan con la condición establecida, deben cambiarse a "soltero" en estado_civil.

Elaboración: INEC

Para la corrección de datos, previamente se deben consultar y validar los criterios (descripción) con las unidades de análisis y con la institución fuente.

2.2.7 Unicidad

Aquí se revisa que los registros sean únicos a nivel de unidad de análisis. Es decir, se identifican y remueven las unidades duplicadas (Economic Commission for Europe, 2012). Para eso se requiere de un análisis individual por cada registro administrativo, ya que en muchos casos se considerará más de una variable participante, sobre todo en los casos que registran datos en diferente periodo.

En suma, de esta macro actividad salen casos considerados como únicos, mientras que los datos identificados como no únicos pueden tener los siguientes tratamientos: 1) Identificar los casos duplicados y seleccionar aquel que deba ser considerado como único o, 2) Aplicar algún método de integración: suma, máximo, mínimo, promedio, etc., con el cual se deje como único; para mantener la completitud e integridad de los datos, los casos duplicados se deben señalar y no eliminar.

2.2.8 Integración

La integración cumple dos propósitos: 1) construcción de tablas históricas y, 2) generación de registros temáticos.

- **CONSTRUCCIÓN DE TABLAS HISTÓRICAS (INTEGRACIÓN LONGITUDINAL)**

Tiene como objetivo la generación de tablas de datos históricas, mismas que contienen información de los eventos/hechos que han ocurrido en el tiempo sobre un determinado ente; esta tabla almacenará el "curso de vida" (cambios en la vida de la población) de los entes (persona o empresa).

Estas tablas se alimentan con los registros nuevos, modificados, eliminados y reingresados, identificados en cada periodo (a través del "número de control"). Para diferenciar los registros que se añadan en cada periodo, esta tabla contendrá variables adicionales como: fecha corte, fecha fin y, estado (Ver Tabla 6).

- **La fecha corte:** Fecha en la cual la fuente hizo el corte de los datos.
- **La fecha fin:** Fecha hasta la cual están vigente el dato del ente; para los casos que no han sufrido cambios, por defecto se registra la fecha 2999/12/31, pero cuando haya una modificación, la "fecha fin" se reemplazará con la "fecha de corte" del nuevo dato.
- **Estado,** almacena 4 categorías: 1) Nuevos. Casos que no emparejan con la tabla de los periodos anteriores, 2) Modificados. Casos que emparejan con la tabla de los periodos anteriores, pero tienen diferente "Número de Control" (con cambio de dato en alguna variable), 3) Eliminados. Casos que emparejan, con periodos anteriores pero no con la tabla del periodo inmediatamente anterior y, 4) Reingresos. Aquellos casos que emparejan, pero no con la tabla del periodo inmediatamente anterior sino con periodos anteriores, y tenían la categoría de Eliminados (nuevamente aparecen en la tabla de datos).

Tabla 6: Construcción de Registro Histórico T_n y T_{n-1}

ID	Variables	fecha_corte	fecha_fin	Estado
1	v1, …, vn	2015/12/31	2999/12/31	1
2	v1, …, vn	2015/12/31	2999/12/31	1
3	v1, …, vn	2015/12/31	2999/12/31	1
4	v1, …, vn	2015/12/31	2999/12/31	1
5	v1, …, vn	2015/12/31	**2019/03/30**	1
6	v1, …, vn	2015/12/31	**2019/03/30**	1
7	v1, …, vn	2015/12/31	2999/12/31	1
5	v1, …, vn	**2019/03/30**	2020/12/31	2
6	v1, …, vn	**2019/03/30**	**2999/12/31**	2
8	v1, …, vn	**2019/03/30**	**2999/12/31**	1
5	v1, …, vn	2020/12/31	2999/12/31	2

Nota: Si en el siguiente periodo surgen cambios sobre los mismos casos (5 y 6), la "fecha fin" debe ser modificada únicamente cuando sea igual a 2999/12/31

Elaboración: INEC

La Tabla 6 muestra un ejemplo de tabla de datos histórica; esta contiene la(s) variable(s) de identificación ID y las n variables (v1,…, vn) de uso estadístico. Aquí se puede ver que en t_0 (fecha_corte = 2015/12/31), los casos del 1 al 7 ingresaron como nuevos (estado = 1). Por otro lado, en t_1 (fecha_corte = 2019/03/30) llegan los casos 5, 6 (con algún cambio) y 8 (nuevo), mismos que deben ser añadidos a la tabla de datos anterior considerando los valores de fecha_corte = 2019/03/30, fecha_fin = 2999/12/31 y, estado de acuerdo a la condición que aplique. Además, para los casos 5 y 6, que sufrieron modificación en sus datos con respecto al periodo anterior (t_0), se debe modificar la fecha_fin del periodo anterior a 2019/03/30. Finalmente se puede ver que el t_2 llegó el caso 5 con alguna modificación, al cual se le aplica el mismo procedimiento, pero en la columna fecha_fin se modifica únicamente aquel que tenía fecha abierta (2999/12/31) por 2020/12/31.

- **CONSTRUCCIÓN DE TABLAS DE DATOS TEMÁTICAS (INTEGRACIÓN TRANSVERSAL)**

Mediante la integración transversal se incorporarán nuevas variables y/o nuevos casos de otros registros; con lo cual se enriquece el análisis de cierta temática, y se añaden nuevos casos se cubren o corrigen problemas de cobertura/sub-registro. Aquí es importante considerar los siguientes aspectos: los datos deben tener la misma fecha de corte (por ejemplo, personas fallecidas en el año 2020), las tablas de datos deben contar con variables de vinculación, mismas que deben pertenecer a la misma población de análisis y, cuando se trate de corregir la cobertura de un registro temático, estos deben tener las mismas variables (conceptualmente ser lo mismo).

Para tener la trazabilidad de los datos, las nuevas variables que se añadan deben poseer un sufijo que indique la fuente desde donde proviene, así también, cuando se integren variables comunes desde varias fuentes, es necesario contar con una matriz de priorización, misma que definirá cuáles son las fuentes primarias y las secundarias. Para este tipo de integración, las variables independientemente del nombre, conceptualmente deben ser las mismas.De esta macro actividad se construyen tablas históricas o tablas temáticas.

2.2.9 Seudonimización

La seudonimización es un conjunto de técnicas que permiten reducir la probabilidad de identificación de los entes (personas o empresas) y ayuda a los analistas a cumplir con tareas de análisis de datos sin vulnerar la confidencialidad estadística. Es

importante indicar que la seudonimización es distinta a la anonimización, ya que esta última es un proceso irreversible de reidentificación.

Con base en el dictamen sobre técnicas de protección de los datos del "grupo de trabajo de la Unión Europea" (Grupo de trabajo sobre protección de datos, 2014), la seudonimización presenta las siguientes bondades:

- **Singularización.** Los datos pueden particularizarse, haciendo posible extraer de un conjunto de datos los registros de interés específico, ya que un individuo queda identificado por un atributo único que es el resultado de la seudonimización.
- **Vinculabilidad.** Cuando se usa el mismo código seudonimizado para referirse a un individuo en distintas fuentes (censos, encuestas y registros administrativos), es posible su vinculabilidad con otros registros.
- **Inferencia.** Se puede inferir la identidad real de un individuo a través de la combinación de variables (seudoidentificadores) de la misma fuente u otras fuentes que usen el mismo atributo seudonimizado (datos vinculados), o bien en el caso de que los seudónimos sean autodescriptivos y no enmascaren adecuadamente la identidad del interesado.

Aquí se reemplazan las variables de identificación directa (cédula, RUC, etc.) por un código de identificación único y confidencial generado por el área o institución encargada de la seudonimización. Además de la eliminación de las variables de identificación auxiliares y de contacto, tales como: nombres y apellidos, razón social, nombre comercial, teléfono, e-mail, entre otras.; la técnica que se propone es la "Descomposición en tokens", que consiste en el reemplazo de los números de identificación original por valores que no tienen utilidad para las personas que accedan a los datos de forma fraudulenta (Vazquez & De Miguel, 2017).

Es necesario construir previamente la "Tabla de correspondencia", misma que almacene las relaciones entre los datos de identificación individual provenientes de distintas fuentes, y los códigos de identificación seudonimizados generados (encargadas de la seudonimización).

Entre las principales variables (para empresas u otras, se debe establecer su símil) que contendrá esta tabla de datos son: identificación directa de un individuo, código seudonimizado creado, fecha de ingreso a la tabla de datos.

Ilustración 12: Seudonimización de datos de identificación

CÉDULA	NOMBRES	FECHA DE NACIMIENTO	LUGAR DE NACIMIENTO	SEXO	CÓDIGO SEUDONIMIZADO
1717593626	Ángel Guano	01/06/1900	03-01-01	1	1110001110
1717593621	Luis Días	03/01/1987	17-01-50	1	1110001111
1300112567	Damián López	11/12/1990	13-50-03	1	1110001112
2111456788	Karla Valdivieso	05/12/1988	21-50-02	2	1110001113
1523675432	Lorena Garcés	08/05/1985	15-50-01	2	1110001114

TABLA DE CORRESPONDENCIA

Se reemplaza las variables de identificación por un código único propio del INEC
Elaboración: INEC

Se deben crear tablas de correspondencias de acuerdo a las unidades de observación, entre ellas: personas, empresas y otras.

El código seudonimizado estará conformado con una estructura lógica tal que permita relacionar y diferenciar a las personas de las empresas, para así tener una relación entre los códigos, siguiendo la misma lógica del RUC de una persona natural; con esto se evitará generar códigos distintos para un mismo individuo o empresa.

Es importante indicar que se generarán códigos únicamente para aquellos casos cuya identidad (Verifica_Id) sean válidas válidos verdaderos y válidas falsas.

Para los casos en que se requiera revertir la seudonimización por un tema específico, se debe tomar en cuenta lo siguiente: no puede revertirse más del 5% del total de registros de la base seudonimizada, la reversión se realiza en ambientes controlados que incluyan auditorias y cuotas asignadas y, la reversión debe ser sólo para casos que sean atípicos y distorsionen los análisis.

Es importante aclarar que la 'Tabla de correspondencias' no almacena todos los datos de identificación de la tabla que se seudonimizará, pues esta tendrá únicamente las principales variables de identificación y su correspondiente código seudonimizado. Como resultado de esta macro actividad se obtienen datos seudonimizados para proveer a las unidades encargadas del análisis.

3. Conclusiones y recomendaciones

Los datos captados en los registros administrativos por las instituciones públicas y privadas, en su mayoría no tienen propósitos estadísticos, por ello es necesario que estos datos pasen por un proceso de transformación para ser utilizados con propósitos estadísticos.

Se establecen nueve macro actividades para transformar registros administrativos en registros estadísticos: perfilamiento, corrección, estandarización, precisión, identificación de cambios, coherencia, unicidad, Integración y, seudonimización, mismas que permiten seguir un flujo secuencial, que pueden llevar a la automatización de todo el proceso.

Dado que los registros administrativos son dinámicos en el tiempo, su volumen crece constantemente. Por lo tanto,para su procesamiento se deben tomar en cuenta el uso de tecnologías que estén a la vanguardia para el manejo de grandes volúmenes de datos (big data).

Para la recopilación frecuente y segura de datos desde las fuentes administrativas (instituciones), se recomienda considerar distintos escenarios que se acoplen a la tecnología que disponga la institución que provee los registros, ya que cada institución tendrá distinta infraestructura tecnológica para la transferencia de datos.

Se recomienda también que las instituciones o áreas que manejan registros administrativos con fines estadísticos apliquen procesos que impidan revelar la identidad de un ente (seudonimización), con lo cual se garantiza la vinculación entre varios registros estadísticos con el fin de robustecer el análisis, así también, se evita ocasionar daño o violación de la privacidad, y se genera un ambiente de confianza con los proveedores de datos.

4. Bibliografía

Asamblea Nacional de Ecuador. (26 de Mayo de 2021). Ley Orgánica de Protección de Datos Personales. Quito. Recuperado el 09 de Junio de 2021, de https://www.asambleanacional.gob.ec/sites/default/files/private/asambleanacional/filesasambleanacionalnameuid-29/Leyes%202013-2017/920-lmoreno/ro-459-5to-sup-26-05-2021.pdf

Bravo, D. M. (2015). *UF1275 - Selección, instalación, configuración y administración de los serrvidores de transferencia de archivos.* España: Elearning s.l.

Cáceres, E. (2008). *www.facso.unsj.edu.ar.* Recuperado el 08 de Junio de 2021, de http://www.facso.unsj.edu.ar/catedras/ciencias-economicas/sistemas-de-informacion-I/documentos/tabain.pdf

CONAGE. (2010). *https://iedg.sni.gob.ec.* Recuperado el 08 de 03 de 2021, de https://iedg.sni.gob.ec/geoportal-iedg/documentos/perfil_ecuatoriano_metadatos_pem.pdf

DANE. (2009). Plan de Fortalecimiento de Registros Administrativos. Colombia.

Economic Commission for Europe. (2012). Model of transformation of administrative data to statistical. *UNECE- Eurostat Expert Group Meeting on Censuses Using Registers* (pág. 3). Geneva: UNECE- Eurostat.

Grupo de trabajo sobre protección de datos. (2014). Técnicas de anonimización. *Dictamen 05/2014 sobre técnicas de anonimización*, 42.

INEC. (09 de Octubre de 2014). *Norma técnica para la producción estadística básica.* Quito.

INEC. (2016). *Modelo de Producción Estadística del Ecuador.* Quito: INEC.

INEC. (2019). Hacia un sistema de estadísticas basadas en registros administrativos: una propuesta metodológica para evaluar registros administrativos. *REVISTA DE ESTADÍSTICA Y METODOLOGÍAS*, 7-20.

INEGI. (2012). *Proceso estándar para el aprovechamiento de registros administrativos.* México: INEGI.

Oracle. (2022). *CREATE DATABASE LINK*. Obtenido de CREATE DATABASE LINK: https://docs.oracle.com/cd/B19306_01/server.102/b14200/statements_5005.htm

Oracle. (2022). *CREATE MATERIALIZED VIEW*. Obtenido de CREATE MATERIALIZED VIEW: https://docs.oracle.com/cd/B19306_01/server.102/b14200/statements_6002.htm

RAE. (2021). *RAE*. Obtenido de https://dle.rae.es/internet

United Nations Economic Commission for Europe. (2007). *Register-based Statistics in the Nordic Countries - Review of best practices with focus on population and social statistics.* New York.

Vazquez , S., & De Miguel, J. (29 de Enero de 2017). *www.confilegal.com.* Recuperado el 15 de Agosto de 2018, de https://confilegal.com/20170129-la-importancia-del-seudonimizacion-en-el-nuevo-reglamento-de-proteccion-de-datos/

Wallgren, A., & Wallgren, B. (2007). *"Register-based Statistics" Administrative Data for Statistical Purposes.* Sweden: Wiley.

Wallgren, A., & Wallgren, B. (2012). *"Estadísticas basadas en registros" Aprovechamiento estadístico de datos administrativos.* Mexico: INEGI.

5. Apéndice: Términos y Definiciones

Institución Fuente: Institución u organismo encargado del levantamiento y administración del registro administrativo.

Método de transferencia de información: Acto de distribuir o proveer acceso a información almacenada digitalmente, utilizando un canal de comunicación de un sistema a otro.

Medios de transferencia: Canal que permite transmitir información entre dos terminales de un sistema.

Estructura o modo de transferencia de información: Recursos tecnológicos que permiten transferir información.

Número de control: Código (número) que se genera mediante operaciones matemáticas entre el código ASCII y la posición de cada letra, que sirve para identificar los cambios producidos por la fuente (institución proveedora) en el tiempo $t_{(n)}$ y $t_{(n-1)}$ de un mismo objeto.

Registro administrativo: Registro usado para fines administrativos en un sistema de información administrativa. Contendrá todos los objetos para administrar, sus objetos serán identificables y sus variables se usarán para propósitos administrativos (Wallgren & Wallgren, 2012).

Registro estadístico: Registro procesado para propósitos estadísticos. Se crean mediante el procesamiento de registros administrativos de modo que los conjuntos de objetos y las variables satisfagan necesidades estadísticas (Wallgren & Wallgren, 2012).

Seudonimización: Tratamiento de datos personales de manera que ya no pueda atribuirse a un titular sin utilizar información adicional, siempre que dicha información adicional, figure por separado y esté sujeta a medidas técnicas y organizativas para garantizar que los datos personales no se atribuyan a una persona física identificada o identificable (Asamblea Nacional de Ecuador, 2021).

Tabla de datos: Archivo contenedor de datos, formado por filas y columnas. Todos los registros/filas tienen los mismos campos/columnas. La igualdad de forma de los registros en relación a los mismos campos, es lo que forma la estructura de un registro genérico (Cáceres, 2008).

Recopilación: Acción de diligenciar y compilar en los institutos de estadística los datos provenientes de distintas fuentes.

@ecuadorencifras

@ecuadorencifras

@InecEcuador

t.me/ecuadorencifras

INEC/Ecuador

INECEcuador

INEC Ecuador

WWW.ECUADORENCIFRAS.GOB.EC