

**Cooperación Técnica No Reembolsable
No. ATN/OC-14340-RG – BID**

***“Proyecto Estadística de Población e Inmuebles a partir
del uso de registros administrativos oficiales en la
Comunidad Andina”***

**MARCO CONCEPTUAL Y METODOLÓGICO QUE SUSTENTA EL DISEÑO,
DESARROLLO E IMPLEMENTACIÓN DE UN SISTEMA INTEGRADO DE
REGISTROS ESTADÍSTICOS DE POBLACIÓN E INMUEBLES**

Federico Segui Stagno

Diciembre de 2016

Tabla de contenido

1.	Introducción	4
2.	Base conceptual	5
3.	Modelos de referencia	8
4.	Experiencias internacionales.....	13
5.	Diagnóstico subregional	20
6.	La gestión por procesos como modelo de gestión de la producción de estadísticas basadas en registros administrativos.....	20
6.1.	Modelos de referencia	21
6.1.1.	Marco de referencia Arquitectura Común para la Producción Estadística (CSPA) 22	
6.1.2.	Modelo Genérico de Procesos Misionales de Producción Estadística (GSBPM). 23	
6.1.3.	Modelo genérico de actividades de las ONE (GAMSO).....	25
6.1.4.	Modelo Genérico de Información Estadística (GSIM).....	28
6.2.	GSRBPM – Modelo Genérico de Procesos de Producción de Registros Estadísticos (que forman parte del sistema integrado de registros estadísticos)	29
6.3.	Diseño de procesos	43
7.	Aspectos organizacionales	50
8.	Marco conceptual y metodológico común	52
8.1.	Elementos del sistema	53
8.2.	Categorías e interrelaciones de Registros Estadísticos	55
8.3.	Transformación e integración de RA en RE.....	56
8.3.1.	Controles de consistencia y calidad y depuración de datos	57
8.3.2.	Variables del sistema.....	58
8.3.3.	Unión de registros	67
8.4.	Resumen de pasos para la implementación de la metodología	75
8.5.	Adaptaciones a cada país	76
8.6.	Plan de implementación estándar	76
9.	Metadatos, herramientas e instrumentos para la documentación.....	77
10.	Modelo de evaluación de la calidad del sistema integrado de registros estadísticos	78
11.	Metodología para la gestión de inventarios nacionales de RA	86
12.	Sistema integrado de registros estadísticos – SIREPI.....	86
12.1.	Características básicas del SIREPI.....	87
12.2.	Bases para el diseño del SIREPI	88
13.	Glosario	93

14. Bibliografía	99
Anexo I – Síntesis del diagnóstico subregional	102
Anexo I.1 – Observaciones sobre los instrumentos de evaluación utilizados	124
Anexo I.2 – Ficha de requerimientos de información a cada país	131

1. Introducción

Las demandas de información con mayor nivel de desagregación geográfica, cobertura temática y periodicidad han ido en aumento en los últimos años.

Los nuevos requerimientos de datos oportunos provenientes de los sistemas sub-nacionales de estadística y los gobiernos locales han obligado a los INE a buscar fuentes de información que permitan generar estadísticas a un mayor nivel de desagregación territorial y con menor costo.

La generación de indicadores para medir los Objetivos de Desarrollo Sostenible es una nueva demanda de información que en muchos casos es muy difícil de satisfacer a través de los métodos tradicionales de producción estadística.

En todo esto los registros administrativos cobran un rol preponderante. Esta información administrativa ya ha sido capturada por las entidades del estado y puede ser utilizada además con fines estadísticos (previo procesamiento y transformación) a bajo costo.

Los registros administrativos se caracterizan en que tienen amplia cobertura de la población objetivo, no han sido concebidos con fines estadísticos, la lógica o procesos que manejan son diferentes a la secuencia de pasos estadísticos, los instrumentos de captura de datos están orientados a seguimiento y control administrativo, la periodicidad puede estar establecida por normativa o no.

La generación de estadísticas a partir de registros administrativos presenta una serie de ventajas:

- 1) Bajo costo.
- 2) Menor carga para los informantes.
- 3) Evita duplicación de esfuerzos entre entidades públicas.
- 4) Mayor cobertura.
- 5) En general tienen mayor tasa de respuesta.
- 6) No hay errores de muestreo
- 7) Es posible generar estadísticas con mayor desagregación. Estadísticas locales y subpoblaciones.
- 8) Permite fortalecer los sistemas de información nacional, a todo nivel.
- 9) Mejora la calidad de la investigación, pues permite combinar varias fuentes y realizar análisis multidimensional.
- 10) Facilita análisis longitudinales.
- 11) Facilita la construcción de series temporales.
- 12) Fuente de información de los indicadores de los ODS.

Pero también se encuentran ciertas desventajas:

- 1) Diferentes unidades de análisis entre unidades administrativas y las estadísticas.
- 2) Diferencias en definiciones y conceptos. Definiciones de variables.
- 3) Falta de documentación, rigor metodológico.
- 4) Difícil implementar cambios en los instrumentos y procesos de captura de datos de las fuentes administrativas.
- 5) Cambios políticos e institucionales pueden afectar la continuidad de los registros administrativos.
- 6) Falta de identificador común estandarizado para la integración de datos de diferentes fuentes.
- 7) Dependencia de la cooperación interinstitucional y marco legal adecuados.
- 8) Cambio cultural en los INE, resistencia al cambio, se “pierde el control” sobre el proceso de captura de datos.

Según UNECE¹, mediante el uso de métodos tradicionales de recolección de datos, los INE se plantean la siguiente paradoja: "hicimos nuestra pregunta, pero ¿tenemos la respuesta correcta?", en cambio, usando datos provenientes de registros administrativos a los INE se les plantea la siguiente interrogante: "seguramente tenemos la respuesta correcta, pero ¿cómo corresponde a nuestra pregunta?". Los INE están trabajando hacia el mismo objetivo, pero desde un punto de partida diferente.

2. Base conceptual

Los países escandinavos (Suecia, Noruega, Finlandia, Dinamarca, Islandia) han sido los pioneros en el uso de registros administrativos con fines estadísticos. Su vasta experiencia está siendo compartida y replicada o adaptada en varios países de la Unión Europea y el mundo.

La metodología escandinava, en particular el modelo sueco, para la generación de estadísticas a partir de registros administrativos ha sido plasmado en el trabajo de A. Wallgren y B. Wallgren (2012). *Estadísticas basadas en registros. Aprovechamiento estadístico de los registros administrativos*. INEGI.

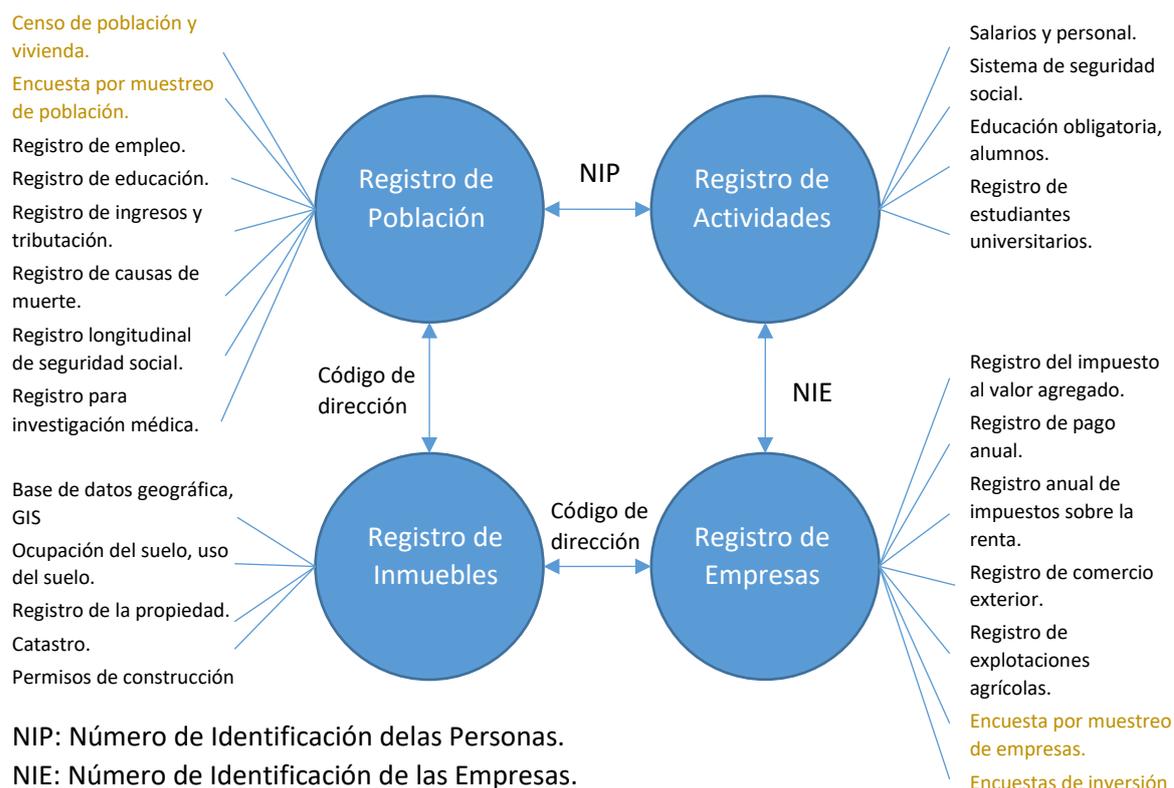
Esta experiencia ha sido ampliamente difundida en la región y se ha tomado como base conceptual para desarrollar el presente marco conceptual y metodológico.

¹ UNECE (2007). *Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics*. Naciones Unidas. Nueva York y Ginebra 2007.

El modelo implementado por la oficina de estadística de Suecia está orientado hacia un sistema estadístico basado en un sistema integrado de registros estadísticos que consta de cuatro registros base (población, empresas, inmuebles y actividades) integrados entre sí.

La siguiente figura ilustra el modelo conceptual genérico de un sistema de registros estadísticos.

Figura 1. Sistema de registros estadísticos por tipo de objeto y campo de estudio.



Fuente: Anders Wallgren, Britt Wallgren. (2012). *Estadísticas basadas en registros. Aprovechamiento estadístico de los registros administrativos*. INEGI.

Actualmente, la oficina de estadística de Suecia cuenta con un sistema de registros conformado por tres registros base (población, empresa e inmuebles)² al igual que el resto de los países nórdicos.

Reviste poca trascendencia el hecho que sean tres o cuatro los registros base, cuando lo relevante es contar con un sistema integrado de registros estadísticos compuesto por ciertos registros base que cuenten con buena cobertura y calidad.

En los países nórdicos y Holanda es obligatorio registrar cualquier cambio de domicilio, lo cual ha facilitado la conformación del registro base de población. En

² Klas Blomqvist and others (2011). *A strategy to improve the register system to store, share and access data and its connections to a generic statistical information model (GSIM)*. Invited paper. Work Session on Statistical Data Editing of the Conference of European Statisticians – UNECE. Ljubljana, Slovenia, 9-11 May 2011.

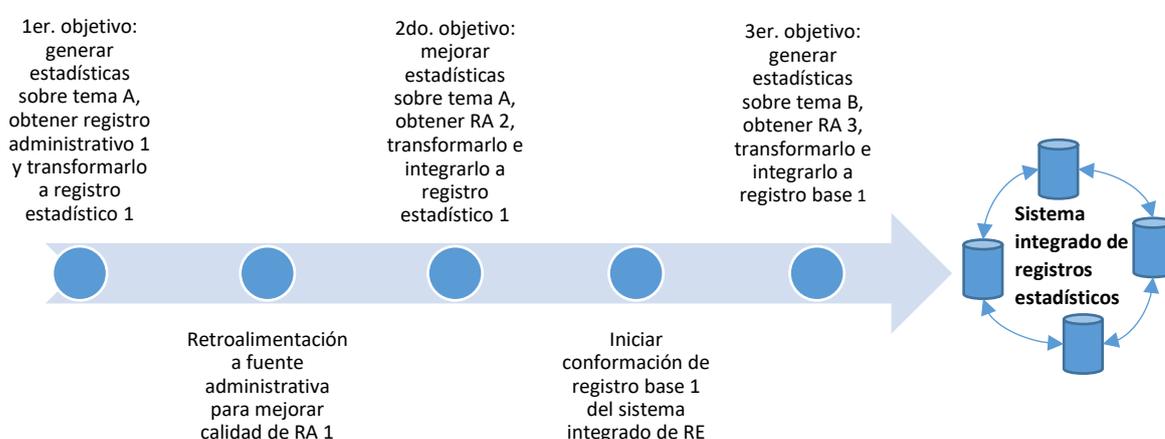
cambio, si no se cuenta con un registro de población residente (población residente en cada municipio, flujos) y un registro de domicilios, se debería utilizar el enfoque de unir múltiples fuentes de datos administrativos existentes.

La oficina de estadística del Reino Unido (ONS) está investigando activamente la factibilidad de realizar un censo basado en registros administrativos sin un registro nacional de población. La experiencia de la ONS también se está considerando en otros países, como Nueva Zelanda entre otros, dadas las similitudes legales y culturales.

En el marco del presente proyecto se ha acotado el alcance del sistema integrado de registros estadísticos a dos registros base (población e inmuebles), a los efectos de facilitar su implementación en los países de la Comunidad Andina; sin perder de vista que el modelo consta de los tres o cuatro registros base descriptos anteriormente.

La premisa es construir paso a paso, empezar por lo que se tiene acceso y de a poco ir conformando los registros base a medida que se generan estadísticas a partir de lo que se dispone y en paralelo se van conformando los registros base. Se deberían definir etapas y trazar objetivos para cada una, según se ilustra en la figura 2. Así también lo recomienda Naciones Unidas³, avanzar paso a paso a partir de lo que se tiene actualmente.

Figura 2. Esquema del plan de acción para la conformación del sistema integrado de registros estadísticos.



Fuente: elaboración propia del consultor.

³ UNECE (2007). *Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics*. Naciones Unidas. Nueva York y Ginebra 2007.

3. Modelos de referencia

En el capítulo anterior se presentó la base conceptual tomando como referencia el modelo nórdico, en especial el modelo sueco, para la creación y mantenimiento de un sistema integrado de registros estadísticos. Sin embargo, existen otros modelos para crear y mantener registros estadísticos implementados en diferentes países y para diversas temáticas⁴.

Las fuentes administrativas disponibles difieren bastante de un país a otro, por lo tanto es difícil adoptar un modelo tal cual o apoyarse en estándares internacionales incipientes.

Los modelos que se presentan a continuación proveen ideas que pueden ser adaptadas a la realidad de cada país, incluso en combinación con la base conceptual planteada al inicio, sustentada en el modelo nórdico.

De hecho, la metodología planteada en el presente documento toma varias de estas ideas y las combina para crear un modelo adaptado al contexto subregional y particularidades de cada país.

a) Combinación de múltiples fuentes

El Reino Unido ha implementado un modelo para crear y mantener el registro estadístico de empresas.

La figura 3 muestra el modelo con el registro estadístico al centro rodeado de las fuentes de datos y lo presenta como la herramienta para combinar y reconciliar los datos provenientes de varias fuentes. Además, utiliza registros “satélite o asociados” y una mezcla de fuentes de datos estadísticos y administrativos.

⁴ UNECE (2011). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. Naciones Unidas. Nueva York y Ginebra, 2011.

Figura 3. Modelo simplificado de las fuentes del registro estadístico de empresas en el Reino Unido.



Fuente: UNECE (2011). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. Naciones Unidas. Nueva York y Ginebra, 2011.

b) Registros administrativos centralizados

Los gobiernos siempre han estado preocupados en lograr mayor eficiencia en la administración pública. En este sentido, se han creado registros administrativos centralizados que proveen un canal único por medio del cual los individuos que forman parte del registro pueden interactuar con las diferentes entidades del gobierno, reduciendo duplicidades y los tiempos de los procedimientos administrativos.

Este tipo de registros administrativos centralizados tiene enormes ventajas tanto administrativas como en su uso estadístico. En los casos que los individuos (personas o empresas) necesiten actualizar información en el registro (por ejemplo, cambios de domicilio) lo deben hacer sólo una vez en el registro central, luego esta información es compartida por todas las entidades que forman parte del registro.

En cuanto al uso con fines estadístico, el registro administrativo centralizado permite reducir los tiempos y errores de unión de registros y conciliación de datos de diferentes fuentes. Es importante para los INE tener algún tipo de injerencia en el diseño y/o gestión del registro administrativo, para asegurarse

que cumple con los requerimientos estadísticos (unidades, clasificaciones, definiciones), tanto como sea posible.

El registro australiano de empresas implementado por la oficina de impuestos de Australia es un buen ejemplo de este tipo de modelo. Este registro central de empresas es mantenido en cooperación con la oficina de estadística (ABS). Otro ejemplo más cercano es el registro central de empresas de la oficina de impuestos (DGI) de Uruguay que lo mantiene en colaboración con el INE.

c) Concentrador para el intercambio de datos

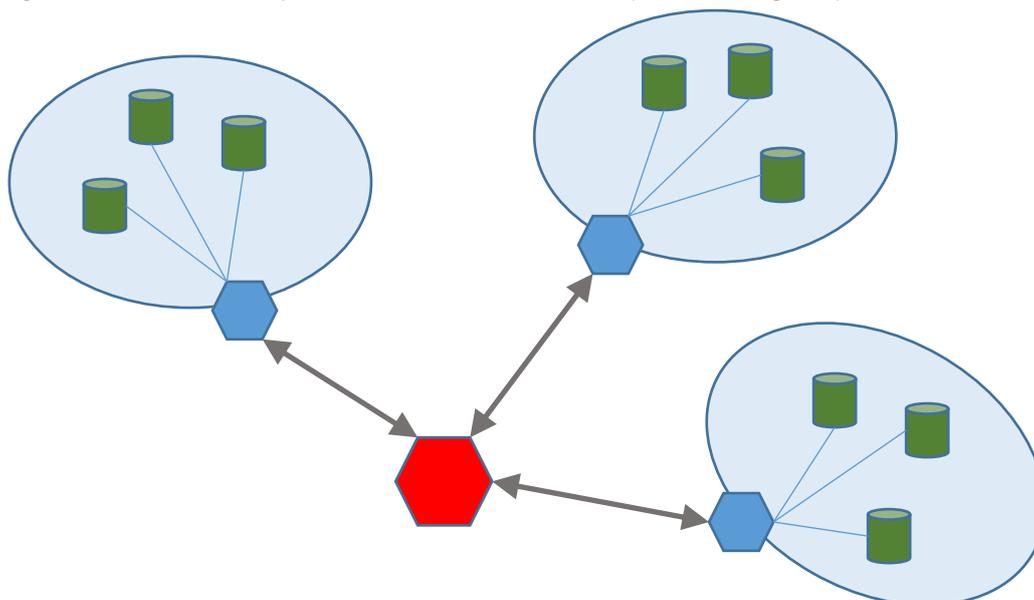
El concentrador para el intercambio de datos es una variante del modelo basado en un solo registro administrativo centralizado. En este caso, el elemento central no es un registro completo, más bien es una herramienta para encontrar y unir datos de varias fuentes administrativas.

La figura 4 ilustra como este núcleo central (hexágono rojo), por medio de información básica de identificación y ubicación de los datos, abre una puerta para que los datos de diferentes agencias pueden ser compartidos dentro del gobierno.

Los óvalos azules representan las diferentes entidades gubernamentales, cada una con diversos contenedores de datos (cilindros verdes). Cada uno de estos contenedores de datos está vinculado a un portal (hexágonos azules) que controla estrictamente qué datos pueden pasar y para quién.

Estos portales están a su vez vinculados al núcleo central que contiene los metadatos necesarios para realizar la búsqueda y unión con los contenedores de datos vinculados. De esta manera, los usuarios de las organizaciones participantes pueden hacer consultas a través del concentrador central y recibir datos alojados en contenedores de datos de otras organizaciones (siempre y cuando tengan permisos de acceso).

Figura 4. Concentrador para el intercambio de datos (data-sharing hub).



Fuente: UNECE (2011). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. Naciones Unidas. Nueva York y Ginebra, 2011.

d) Uso de registros satélite

En este caso los registros administrativos son organizados en registros específicos por cada fuente que cumplen ciertos criterios y están vinculados al registro estadístico.

Los registros satélite o asociados cumplen las siguientes condiciones:

- No son una parte integral de un registro estadístico, pero pueden ser vinculados a éste.
- Su alcance es más limitado que un registro estadístico, pero dentro de su alcance pueden tener mayor cobertura de unidades y/o variables.
- Contienen una o más variables que no se encuentran en el registro estadístico. Estas variables pueden ser utilizadas con fines de estratificación.
- Las bases de datos donde se almacenan los resultados de las encuestas por muestreo no son registros satélite.

Los registros satélite son utilizados para incorporar datos administrativos que son relevantes sólo para un subconjunto de unidades de un registro estadístico. Son creados a partir de información administrativa, encuestas estadísticas o por una combinación de ambas.

Este tipo de registros pueden agregar valor a los registros estadísticos, mejorando la cobertura de la población objetivo, o aumentando la cantidad de variables utilizadas para el análisis estadístico y la estratificación, lo cual

lleva a incrementar la eficiencia de la muestra (mejorando la calidad de las variables de estratificación).

Por ejemplo, en el caso de datos administrativos de empresas, el alcance del registro satélite puede estar determinado por la actividad económica, tamaño (personal ocupado, volumen de ventas) o características específicas (empresas con mayor crecimiento en los últimos años, por ejemplo).

Figura 5. Relación entre un registro satélite y un registro estadístico.



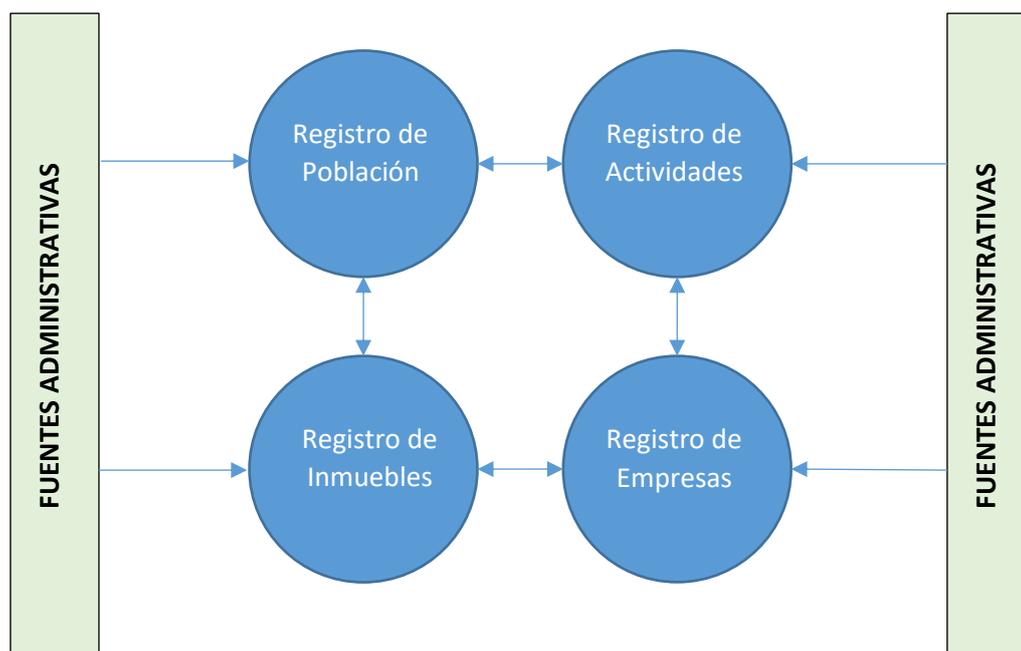
Fuente: UNECE (2011). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. Naciones Unidas. Nueva York y Ginebra, 2011.

e) Sistema integrado de registros estadísticos

Este modelo ya ha sido discutido en profundidad en el capítulo anterior, pero se vuelve a mencionar aquí para incluirlo dentro de la lista de modelos de referencia disponibles.

La principal diferencia con los otros modelos es que varios registros estadísticos vinculados (tres o cuatro registros estadísticos base) se crean utilizando una amplia gama de registros administrativos.

Figura 6. Sistema integrado de registros estadísticos de los países nórdicos.



Fuente: UNECE (2011). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. Naciones Unidas. Nueva York y Ginebra, 2011.

4. Experiencias internacionales

Dinamarca fue el primer país del mundo en realizar un censo completamente basado en registros administrativos, en 1981, seguido de Finlandia en 1990. Noruega y Suecia comenzaron en 1980 a realizar sus censos parcialmente basados en registros administrativos y en 2011 lo hicieron por primera vez utilizando únicamente registros administrativos.

La principal diferencia entre los países nórdicos es que Dinamarca y Finlandia tienen registros de viviendas desde 1980, mientras que Noruega y Suecia los han establecido en 2001 y 2009 respectivamente.

La siguiente tabla muestra la historia resumida sobre el uso estadístico de los registros administrativos en los países nórdicos, en particular en los censos de población y viviendas.

Cuadro 1. Año de establecimiento de los registros e introducción de registros en los censos por país según tipo de registro.

Tipo de Registro	Dinamarca		Finlandia		Noruega		Suecia	
	Año establecido	Usado por 1ª vez en censo	Año establecido	Usado por 1ª vez en censo	Año establecido	Usado por 1ª vez en censo	Año establecido	Usado por 1ª vez en censo
Registro Central de Población	1968	1981	1969	1970	1964	1970	1967	1975
Registro de Empresas	1975	1981	1975	1980	1965	1980	1963	1975
Viviendas	1977	1981	1980	1985	2001	2011	2008?	2011?
Condiciones de los hogares	1977	1981	1980	1985	2001	2011	2008?	2011?
Educación	1971	1981	1970	1975	1970	1980	1985	1990
Empleo	1979	1981	1987	1990	1978	2001	1985	1985
Familia	1968	1981	1978	1980	1964	1980	1960	1975
Hogar	1968	1981	1970	1975	2001	2011	2011?	2011?
Ingresos	1970	1981	1969	1970	1967	1980	1968	1975
Censo totalmente basado en registros		1981		1990		2011		2011?

Fuente: UNECE (2007). Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics. Naciones Unidas. Nueva York y Ginebra 2007.

Holanda

La oficina de estadística de Holanda cuenta desde el año 1850 con un sistema de registros de población, basados en los registros administrativos de población de los municipios. El registro básico de población forma parte del sistema de registros básicos interconectados, compuesto por otros registros como empresas, inmuebles y direcciones, y vehículos entre otros.

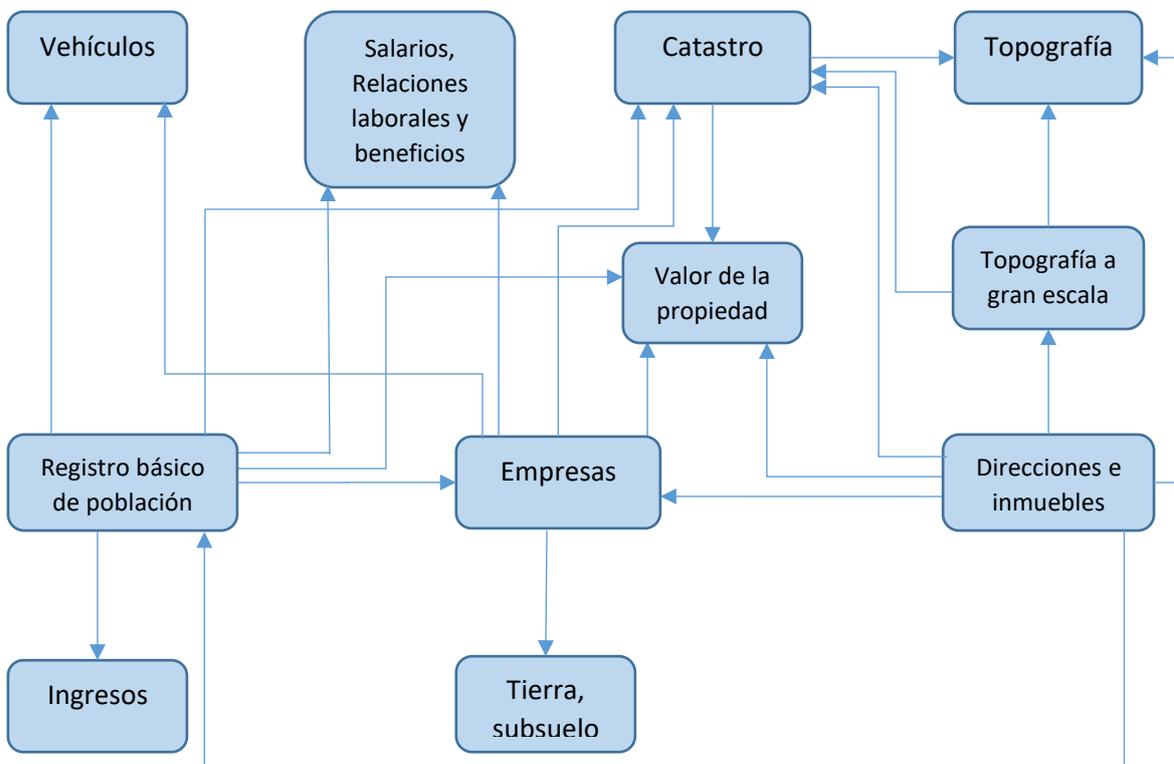
Cada municipalidad está obligada a mantener actualizado un registro de toda la población residente dentro de los límites de la municipalidad. Pero no sólo la población residente en Holanda forma parte del registro de población, sino que también están incluidas personas no residentes pero que tienen alguna relación con alguna agencia del gobierno holandés.

Las principales características del sistema de registros estadísticos sociales (SSD) son:

- Los datos son almacenados centralizadamente en forma estandarizada.
- Los tipos de unidades (personas, inmuebles, hogares, empresas) pueden ser vinculadas fácilmente pues se dispone de claves de unión de registros asignadas.
- La coordinación es crucial para obtener resultados consistentes. La coordinación abarca aspectos organizacionales, técnicos y de contenido.

La siguiente figura representa el sistema de registros estadísticos básicos de la oficina de estadística de Holanda.

Figura 7. Sistema de registros estadísticos básicos de Statistics Netherlands.



Todos los objetos en el sistema de registros tienen un número de identificación único, lo que permite la interconexión de los registros básicos. Por ejemplo, las direcciones en el Registro Básico de Población y en el Registro de Empresas están conectadas al Registro Básico de Direcciones e Inmuebles.

Cuando un residente comunica su domicilio al municipio para ser registrado en el Registro Básico de Población, las autoridades municipales están obligadas a verificar si esta dirección está en el Registro Básico de Direcciones e Inmuebles. Si este es el caso, el número de identificación de esa dirección se registra como la dirección de la persona.

Sin embargo, si la dirección no está en el Registro Básico de Direcciones e Inmuebles, la persona es registrada en esa dirección, pero las autoridades municipales deben reportar la presunta omisión al Registro Básico de Direcciones e Inmuebles.

El departamento municipal responsable de ese registro debe iniciar una investigación sobre la naturaleza de la dirección. Esta investigación puede conducir a tres resultados diferentes. Un resultado es que la dirección es de hecho una omisión en el Registro Básico de Direcciones e Inmuebles, después de lo cual la dirección se añade a ese registro.

Otro resultado posible es que la dirección es una dirección válida, correctamente registrada en el Registro Básico de Direcciones e Inmuebles, pero no está correctamente escrita en los datos de Registro Básico de Población. En ese caso, debe adaptarse la ortografía de la dirección registrada en este último registro.

El último resultado posible es que la dirección no es válida dada las reglas del Registro Básico de Direcciones e Inmuebles. En ese caso, la dirección sigue registrándose en el Registro Básico de Población, pero no se hace ninguna referencia al Registro Básico de Direcciones e Inmuebles.

El Registro Básico de Población duplica ocho elementos del Registro básico de Direcciones e Inmuebles:

- Área pública
- Número de vivienda
- Letra-número de vivienda
- Adicional al número de vivienda
- Código postal
- Ubicación de los residentes
- Código de identificación residencial
- Código de identificación de la dirección

Los datos del registro básico de población se utilizan en otros registros básicos:

- Registro de empresas
- Registro básico de vehículos
- Registro básico de Catastro
- Registro básico de valor de la propiedad

- Registro básico de ingresos
- Salarios, Relaciones Laborales y Beneficios

Los registros básicos se implementan en leyes separadas, bajo la responsabilidad de un ministerio del gabinete. El Ministerio de Infraestructura y Medio Ambiente es responsable de seis de los once registros básicos, el Ministerio del Interior y Relaciones del Reino es responsable de la coordinación dentro del sistema.

Austria

Las estadísticas de población en Austria se basaban tradicionalmente en los censos realizados cada diez años.

Las condiciones básicas para las estadísticas de población cambiaron de dos maneras a partir del censo de 2001: en primer lugar, el Registro Central de Residencia (CRR) basado en procesamiento electrónico de datos, comenzó a funcionar el 1 de marzo de 2002.

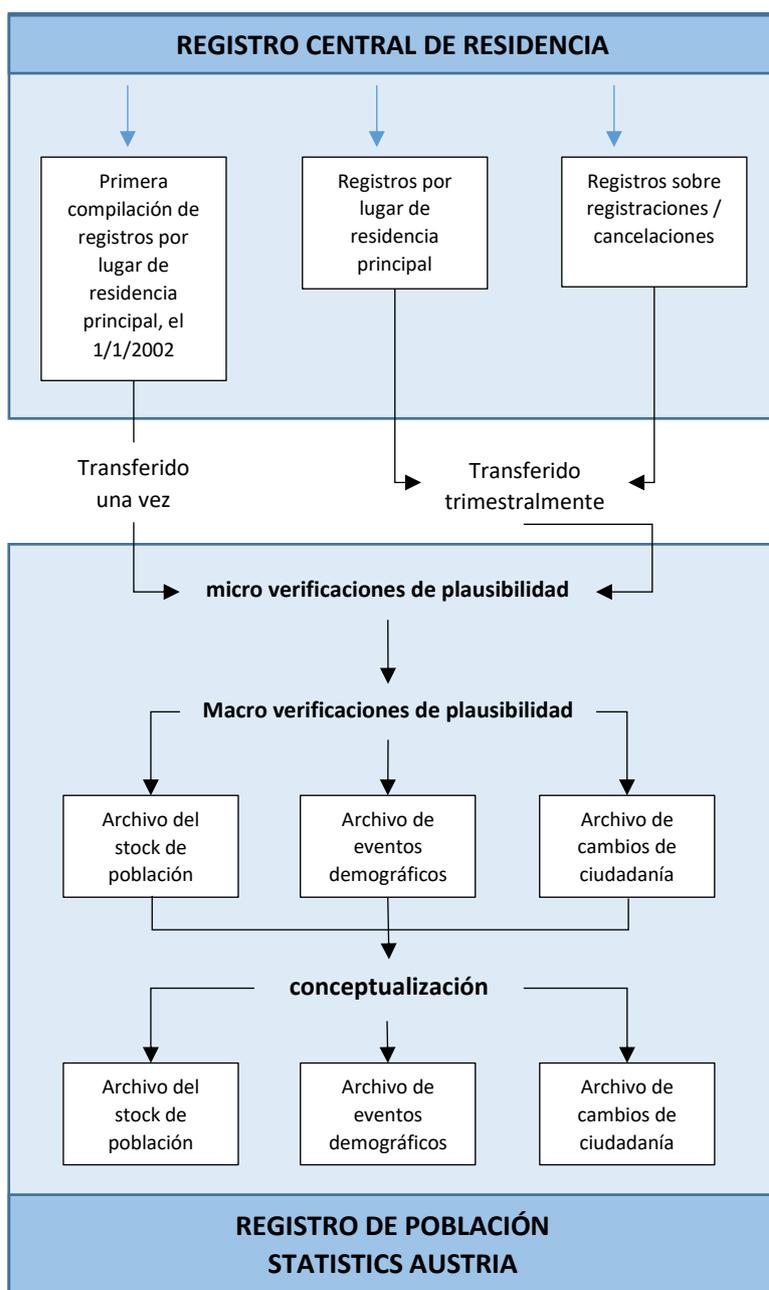
A finales de 2001 se establece que los registros de residencia (altas y modificaciones) deben enviarse a la Oficina de Estadística de Austria en formato electrónico, tal como se establece en la Ley de registro. Además, Statistics Austria recibe trimestralmente un expediente de personas inscritas en el CRR con un domicilio principal (habitual). Las personas en Austria están obligadas a registrarse y actualizar su residencia principal en un plazo de tres días hábiles.

Todos los registros de datos entregados por el CRR contienen el denominado número CRR, que es un identificador único para cada residente en Austria.

En segundo lugar, el Consejo de Ministros de junio de 2000 establecía la voluntad política de no realizar más censos tradicionales de población, sino que el país se disponía a preparar un censo basado en registros administrativos para 2010.

Estas condiciones establecieron las bases para la construcción de un nuevo sistema demográfico. La siguiente figura ilustra el diseño conceptual del nuevo sistema basado en el registro de población denominado "PopReg".

Figura 8. Diseño conceptual del registro de población de la oficina de estadística de Austria.



Nueva Zelanda

Nueva Zelanda no tiene un registro nacional de población y la oficina de estadística neozelandesa tampoco está abogando por uno, sino que mantiene una posición condicional. Es decir, si se creara un registro de población por sus propios méritos y resultara de alta calidad, entonces podría ser la mejor opción a seguir para un censo basado en registros administrativos.

Han considerado esta y varias opciones más para la realización de un censo basado en datos administrativos y las alternativas se resumen en las siguientes dos ⁵:

- Un registro nacional de población como la base para el conteo de población.

La propuesta es crear un registro de población (residencia) donde la función administrativa está estrechamente alineada con la necesidad del censo de contar la población. Este enfoque es la base de ejemplos internacionales exitosos. Sin embargo, en el contexto de Nueva Zelanda, habría que identificar otros beneficios para el gobierno y considerar la privacidad y las cuestiones culturales. Pero no será sino hasta dentro de 20 o 30 años que esperan tener este registro.

- Múltiples fuentes administrativas existentes vinculadas para producir estadísticas de la población.

La vinculación de múltiples fuentes de datos administrativos existentes es prometedora como una solución estadística para producir conteos de población que evitarían la necesidad de entrevistar a toda la población. Este es el enfoque actual de las investigaciones de Statistics New Zealand.

Las dos opciones reflejan la experiencia internacional. La alternativa más directa a un censo basado en registros administrativos sería a través de un registro nacional de población. Entre tanto no exista este registro, el mejor enfoque para generar estadísticas de la población es por medio de la vinculación de múltiples fuentes administrativas de datos de las personas. Obviamente, conlleva una implementación más compleja.

Ambas alternativas deberán trabajar como un sistema integrado de fuentes administrativas que consta de dos registros base. Por un lado, el registro de población, o sino una lista estadísticamente compilada de cada persona que vive en el país, que cumplirían el mismo rol. El otro registro base es el registro nacional de domicilios (una lista de todas las direcciones válidas). Ambos registros base están vinculados a través de las variables que componen el domicilio.

Estos dos registros base (población y domicilios) sirven de espina dorsal de las demás fuentes de datos con información sobre las personas y las viviendas.

Sin un registro nacional de población, será clave la aplicación de métodos estadísticos robustos para la estimación de la población, así como también será necesario realizar una encuesta de cobertura para proveer una validación independiente de los datos administrativos.

⁵ Bycroft, C (2010). *A register-based census: what is the potential for New Zealand?* Wellington. Statistics New Zealand.

5. Diagnóstico subregional

En el Anexo I se incluye la síntesis del diagnóstico subregional que contiene el resumen de los aspectos comunes a los cuatro países de la subregión, principales obstáculos y factores de éxito para la conformación de un sistema integrado de registros estadísticos en cada país.

El informe incluye, además, lecciones aprendidas y recomendaciones para la implementación del sistema.

6. La gestión por procesos como modelo de gestión de la producción de estadísticas basadas en registros administrativos

El modelo de gestión por procesos o gestión basada en procesos está enfocado al cambio operacional de la organización, dejando de lado la operación y organización funcional para pasar a una gestión orientada a los procesos. El siguiente cuadro presenta las principales diferencias entre la gestión centrada en la organización funcional y la gestión orientada a los procesos del INE.

Cuadro 2: cuadro comparativo entre la Gestión centrada en la organización funcional del INE y la Gestión centrada en los procesos del INE.

Gestión centrada en la organización funcional de la ONE	Gestión centrada en los procesos de la ONE
Los funcionarios constituyen la fuente de los problemas de la ONE	Los procesos inadecuados son el problema de gestión de la ONE
Evaluar el desempeño individual	Evaluar el rendimiento de los procesos
Cambiar estructura, personas, funciones, objetivos, etc.	Cambiar los procesos
Orientación a productos y costos	Orientación a resultados, creación de valor y satisfacción de los usuarios internos y externos del SEN
Siempre se puede hallar un funcionario mejor	Siempre se puede mejorar los procesos, "reducir su variabilidad" (W.E. Deming)
Controlar empleados	Desarrollo de las personas, gestión participativa. Auto-control; auto-medida y auto-supervisión
Responsabilidad fragmentada de las tareas de la organización	Responsabilidad colegiada de los resultados finales de la actividad central de la ONE
Optimizar funciones con atención a la eficiencia	Optimizar procesos con atención a la creación de valor

Gestión centrada en la organización funcional de la ONE	Gestión centrada en los procesos de la ONE
Relación jerárquica entre superior y subordinado	Relación proveedor-cliente
Enfrentar y corregir errores	Reducir la variación de los procesos
Consigna a los funcionarios a realizar su trabajo	Comprensión del lugar que ocupa el trabajo de cada funcionario en el proceso y colaboración en el desarrollo de los procesos en que participa

Fuente: Medina, Alejandro y Segui, Federico (2013). *¿Cómo mejorar el desempeño y crear valor público en las oficinas y sistemas estadísticos nacionales en América Latina y el Caribe?* USA: innovacionestadistica.com

En el campo de la gestión por procesos el *cambio* es la meta primaria del diseño de procesos porque *"la capacidad de cambiar es mucho más apreciada que la capacidad de crear en primer lugar"* ⁶.

La metodología BPM (por su sigla en inglés de Business Process Management o Gestión de Procesos de Negocio/Misionales) tiene por objetivo mejorar el desempeño de la organización por medio de la optimización de sus procesos de negocio/misionales. Provee un modelo de gestión flexible que permite a las organizaciones adaptarse rápidamente y con el menor impacto posible a los cambios que cada vez son más frecuentes en la actualidad.

6.1. Modelos de referencia

En este apartado se presenta el estado del arte de los modelos de referencia de gestión de procesos de producción estadística:

- CSPA – Marco de referencia de la Arquitectura común para la producción estadística.
- GSBPM – Modelo Genérico de Procesos Misionales de Producción Estadística.
- GAMS0 – Modelo Genérico de Actividades de las Oficinas de Estadística.
- GSIM – Modelo Genérico de Información Estadística.

⁶ Fingar, P. Smith, H (2003). *Business Process Management. The third wave*. Meghan-Kiffer Press. USA.

6.1.1. Marco de referencia Arquitectura Común para la Producción Estadística (CSPA)

A lo largo del tiempo, las oficinas de estadística han montado estructuras organizativas, procesos de producción e infraestructuras estadísticas y tecnológicas, que en la actualidad se están volviendo insostenibles, o al menos ineficientes⁷.

Los modelos de producción estadística actualmente utilizados por la mayoría de los INE se basan en encuestas por muestreo. Como ya se ha mencionado, cada vez es más necesario que los INE utilicen datos administrativos para producir estadísticas, lo cual implica nuevos procesos y capacidades técnicas/tecnológicas que actualmente no existen en gran parte de los INE.

Los INE han desarrollado sus propios procesos misionales (procesos de negocio) y sistemas de informáticos para producir estadísticas. Sin embargo, cada una de las soluciones tecnológicas ha sido desarrollada para un propósito muy específico, sin considerar el intercambio de información con otras aplicaciones y con limitada capacidad para manejar procesos similares de otros proyectos estadísticos. Esto se debe a que los procesos y las herramientas tecnológicas no fueron diseñados con una visión holística, lo que se denomina “arquitectura accidental” (“surge de la multitud de decisiones individuales de diseño que se producen durante el desarrollo”⁸).

Los INE poco a poco están incorporando enfoques de arquitectura orientados a servicios para mejorar la flexibilidad, robustez y sostenibilidad de sus entornos tecnológicos (tan cambiantes). Están utilizando la arquitectura empresarial para apoyar su visión y cambiar la estrategia.

Una arquitectura empresarial tiene como objetivo crear un entorno que pueda cambiar y apoyar las metas de la organización. Ayuda a eliminar silos, mejora la colaboración entre las diferentes áreas y garantiza que la tecnología esté alineada con las necesidades de la organización.

Algunas oficinas de estadística se han estandarizado dentro de su propia organización, pero no de la misma manera que otros INE.

Es así que el Grupo de Alto Nivel para la Modernización de la Producción y los Servicios Estadísticos (HLG) de UNECE ha desarrollado la Arquitectura Común para la Producción Estadística (CSPA, por su sigla en inglés Common Statistical Production Architecture).

⁷ UNECE (2015). *Common Statistical Production Architecture*. UNECE.

<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture>

⁸ Booch, Grady (2006). *The Accidental Architecture*. IEEE Software. Mayo-Junio 2006.

<https://pdfs.semanticscholar.org/7fdc/6cde5c79f046e4a33e4fb36210abb041b0c6.pdf>

La Arquitectura CSPA proporciona un marco de referencia que incluye principios, procesos y directrices para ayudar a reducir el costo de desarrollo y mantenimiento de procesos y sistemas y para mejorar la capacidad de respuesta del ciclo de desarrollo. Así será más fácil el intercambio y la reutilización de los componentes del proceso, no sólo dentro de las organizaciones, sino entre todas las agencias estadísticas en su conjunto.

La CSPA es la arquitectura empresarial para las entidades productoras de estadísticas oficiales. Es un conjunto de principios y normas comunes acordados para promover una mayor interoperabilidad dentro y entre los diferentes actores que forman parte de la comunidad de productores de estadísticas oficiales, con insumos, procesos, productos y objetivos similares.

La Arquitectura CSPA está separada en cuatro perspectivas:

- Arquitectura empresarial que define lo que hacen las organizaciones del sector y cómo se hace (estadísticas oficiales en nuestro caso);
- Arquitectura de la información que describe la información, sus flujos y usos en todo el sector (estadísticas oficiales), y cómo se maneja esa información;
- Arquitectura de aplicaciones que describe el conjunto de prácticas utilizadas para seleccionar, definir o diseñar componentes de software y sus relaciones, y
- Arquitectura Tecnológica que describe la infraestructura tecnológica de soporte a las otras perspectivas de la arquitectura.

La CSPA se basa en SOA (Service Oriented Architecture o Arquitectura Orientada a Servicios), centrado en los Servicios (servicios estadísticos en este caso). Un servicio es una representación de una actividad de una organización del mundo real con un resultado específico. Es autónomo y puede ser reutilizado por una serie de procesos (ya sea dentro o entre organizaciones estadísticas).

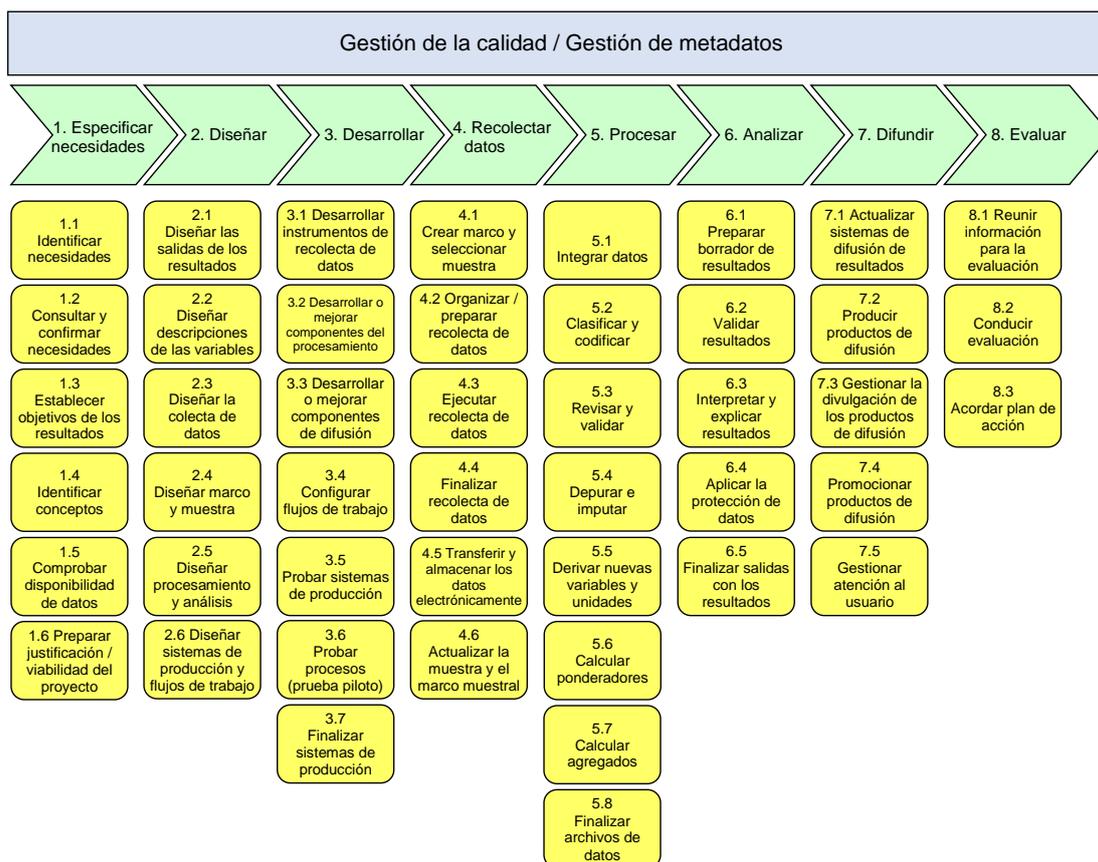
6.1.2. Modelo Genérico de Procesos Misionales de Producción Estadística (GSBPM)

El modelo GSBPM (Generic Statistical Business Process Model) describe y define el conjunto de procesos misionales necesarios para producir estadísticas oficiales. Provee un marco de referencia estándar y terminología armonizada para ayudar a las ONE a modernizar sus procesos de producción de estadísticas.

También puede utilizarse para integrar datos y estándares de metadatos, como plantilla para la documentación del proceso, para armonizar las herramientas tecnológicas que soportan la producción estadística y para proporcionar un marco para la evaluación y mejora de la calidad del proceso.

GSBPM puede ser interpretado y aplicado de manera flexible. No es un modelo rígido en el cual todos los pasos se deben seguir en un estricto orden (si bien su representación gráfica sigue una secuencia lógica, ver figura 9), sino que define los posibles pasos en un proceso de producción estadística y sus interdependencias. Los pasos se pueden ejecutar en diferente orden en diferentes contextos, incluso repitiendo los mismos pasos formando bucles iterativos.

Figura 9. Modelo Genérico de Procesos Misionales de Producción Estadística – GSBPM v.5.0.



Fuente: UNECE (2013). *Generic Statistical Business Process Model – GSBPM v.5.0.*

GSBPM está completamente alineado con los modelos que se describen en los siguientes apartados (GAMSO, GSIM) y proporciona las bases para la implementación de la Arquitectura CSPA.

Figura 10. Relación entre los modelos GSBPM y GSIM.



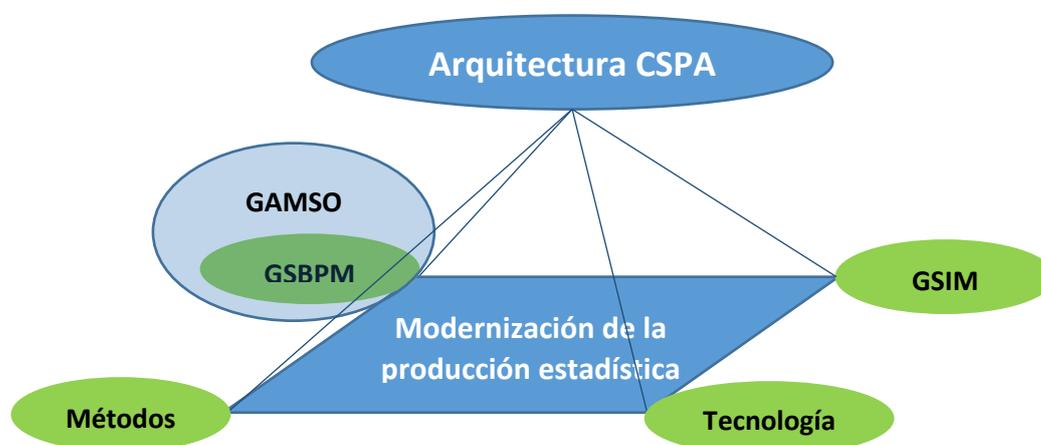
Fuente: UNECE (2013). *Generic Statistical Business Process Model – GSBPM v.5.0.*

6.1.3. Modelo genérico de actividades de las ONE (GAMSO)

El modelo GAMSO describe y define las actividades llevadas a cabo por una típica organización productora de estadísticas. Es una extensión y complemento del modelo GSBPM, pues incorpora actividades adicionales necesarias como soporte de la producción estadística.

La siguiente figura muestra la ubicación del modelo GAMSO en relación a los otros modelos y marcos de referencia necesarios para la modernización de las organizaciones productoras de estadísticas oficiales.

Figura 11. Ubicación del modelo GAMSO en relación a los otros modelos.



Fuente: UNECE (2016). *Generic Activity Model for Statistical Organizations – GAMSO v.1.1.*

El modelo GAMSO está compuesto por tres niveles jerárquicos. El nivel más alto corresponde a: 1) Estrategia y liderazgo, 2) Desarrollo de capacidades, 3) Soporte corporativo y 4) Producción. El segundo nivel son actividades de Estrategia y liderazgo, Desarrollo de capacidades y Soporte corporativo. El tercer nivel proporciona ejemplos de las actividades del segundo nivel. El área de actividades de Producción corresponde al modelo GSBPM, que se ha descrito anteriormente.

El área 1) *Estrategia y liderazgo* se divide en tres actividades:

1.1. Definir la visión.

- Entender las direcciones y contextos nacionales e internacionales.
- Determinar visión, misión y objetivos estratégicos.
- Determinar la propuesta de valor de la organización.
- Determinar y comunicar valores y expectativas.
- Crear interés y conciencia.

1.2. Gobernar y liderar.

- Desarrollar estrategias para alcanzar las metas de la organización.
- Priorizar el acervo de capacidades.
- Priorizar el portafolio de productos y servicios estadísticos.
- Definir el programa estadístico (anual).
- Asignar presupuestos al portafolio de proyectos y programas.
- Construir y mantener la excelencia profesional y estadística interna.
- Asegurar la coordinación y alineación general.
- Definir las políticas generales de la organización.
- Publicar políticas, directrices y documentos normativos.

1.3. Gestionar colaboración y cooperación estratégica.

- Construir y mantener relaciones estratégicas, nacionales e internacionales.
- Construir y mantener la excelencia estadística.
- Avanzar en la colaboración entre organismos nacionales e internacionales.
- Asegurar el apoyo para los productos y servicios estadísticos y el acervo de capacidades.
- Coordinar el sistema estadístico nacional.

2) *Desarrollo de capacidades*

2.1. Planificar mejoras de las capacidades.

- Identificar interrupciones y mejoras en las capacidades.
- Proponer proyectos de mejora de las capacidades.
- Gestionar programas de mejora de las capacidades.

2.2. Desarrollar mejoras de las capacidades.

- Realizar investigaciones de antecedentes.
- Definir requisitos detallados de capacidades.
- Diseñar la solución para la mejora de capacidades.
- Construir / procurar y desplegar soluciones para la mejora de capacidades.

2.3. Monitorear mejoras de las capacidades.

- Mantener la mejora de las capacidades.
- Promover mejora de las capacidades.
- Evaluar la mejora de las capacidades.

2.4. Transferir apoyo de las mejoras de las capacidades.

- Transferir diseño.
- Transferir operaciones.
- Transferir soporte al usuario.

3) *Soporte corporativo*

3.1. Gestionar el desempeño de la organización y la legislación.

- Gestionar el desempeño de la organización.
- Gestionar el cambio y el riesgo.
- Gestionar la legislación y el cumplimiento.

3.2. Gestionar la metodología estadística.

- Marcos muestrales y muestras.
- Edición e imputación.
- Ponderación.
- Estimación.
- Series temporales y ajuste estacional.
- Control divulgación de información confidencial.
- Vinculación de datos.

3.3. Gestionar la calidad.

- Gestionar el marco de la calidad.
- Gestionar las herramientas de aseguramiento de calidad.
- Gestionar la documentación de la calidad.

3.4. Gestionar la información y el conocimiento.

- Gestionar documentos y registros, incluido el archivo y la destrucción.
- Gestionar el conocimiento.
- Gestionar estándares de información y permisos de acceso.
- Gestionar metadatos y datos.

3.5. Gestionar los usuarios.

- Gestionar las comunicaciones y las relaciones con los medios de comunicación.
- Gestionar las consultas con las partes interesadas.
- Gestionar el soporte al usuario.

3.6. Gestionar proveedores de datos.

- Gestionar acuerdos de intercambio de datos.
- Gestionar la transferencia de datos.

3.7. Administrar las finanzas.

- Mantener cuentas (incluyendo activos y pasivos).
 - Gestionar contratos y adquisiciones.
 - Gestionar proveedores de equipos, material de oficina y servicios.
- 3.8. Gestionar recursos humanos.
- Gestionar el desempeño de los empleados.
 - Gestionar y desarrollar habilidades.
 - Gestionar el talento.
 - Gestionar el reclutamiento.
 - Asegurar la planificación de la sucesión.
- 3.9. Gestionar las tecnologías de la información.
- Gestionar los activos y servicios de tecnologías de la información.
 - Gestionar la seguridad informática.
 - Gestionar el cambio tecnológico.
- 3.10. Gestionar edificios y espacio físico.
- Gestionar las necesidades ambientales, mecánicas y eléctricas.
 - Gestionar la disposición del espacio de oficina.
 - Gestionar la distribución de oficinas dentro del espacio.

4) Producción

Las actividades de producción de GAMS0 son aquellas incluidas en el modelo GSBPM v.5.0.

6.1.4. Modelo Genérico de Información Estadística (GSIM)

Este modelo desarrollado por UNECE provee un conjunto de objetos de información estandarizados que son las entradas y salidas en el diseño y producción de estadísticas.

El modelo GSIM permite mejorar la comunicación entre las áreas de la organización, transmisión del conocimiento, automatización del proceso de producción estadística y mejorar la eficiencia de la institución.

GSIM brinda un lenguaje común para describir la información que soporta todo el proceso de producción estadística, desde la identificación de las necesidades de los usuarios hasta la difusión de los productos estadísticos.

El modelo GSBPM ayuda a determinar qué componentes o procesos son necesarios, mientras que el modelo GSIM ayuda a especificar las interfaces como lo ilustra la figura 10.

Es un marco de referencia de objetos de información. Establece definiciones, atributos y relaciones acerca de objetos de información. Está alineado con los estándares de metadatos más relevantes como SDMX y DDI.

6.2. GSRBPM – Modelo Genérico de Procesos de Producción de Registros Estadísticos (que forman parte del sistema integrado de registros estadísticos)

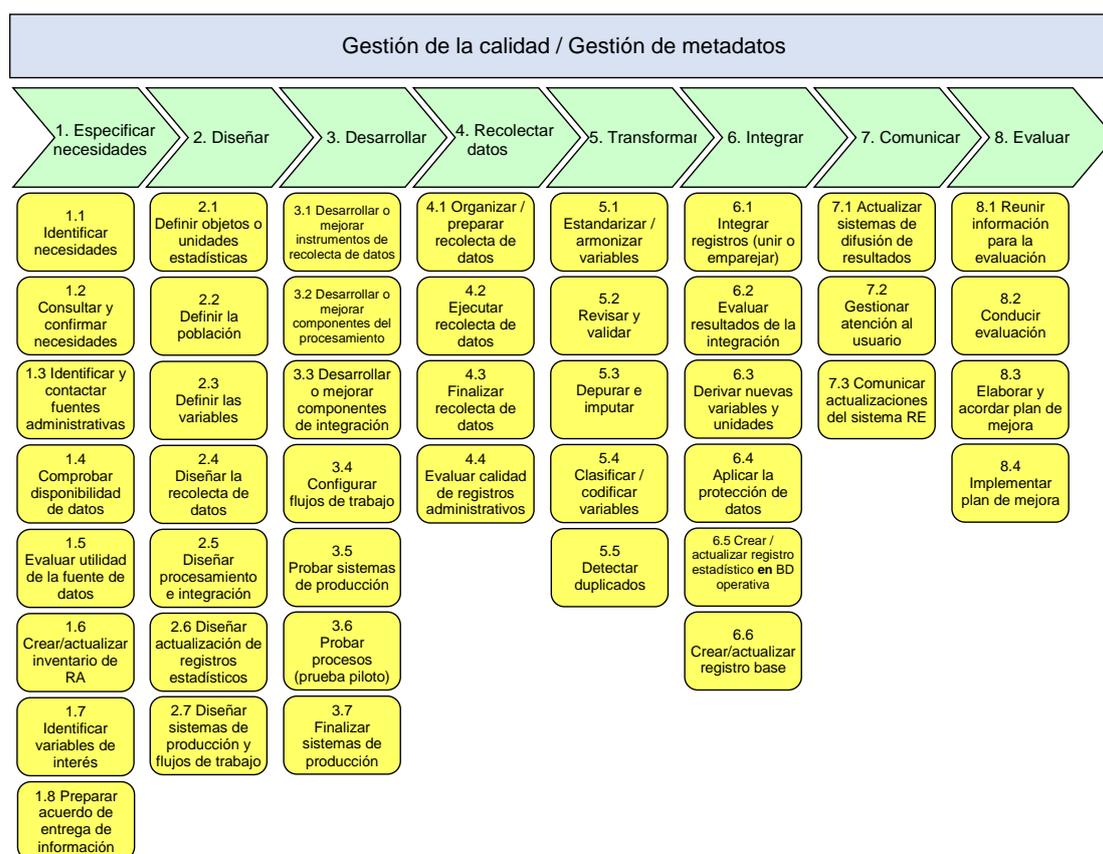
El modelo GSBPM de UNECE está orientado a los procesos de generación de estadísticas a partir en censos, encuestas y registros estadísticos pero tienen un inicio y un fin (incluso cuando la investigación se realiza en forma continua, cada ciclo inicia y termina en un período determinado). En cambio, los procesos de producción (creación y actualización) de registros estadísticos que forman parte de un sistema integrado de registros estadísticos son procesos continuos, independientemente de los usos estadísticos en diferentes proyectos o estudios (obviamente están vinculados a los procesos de los proyectos estadísticos).

Por tanto, se ha adaptado el modelo GSBPM de UNECE y se ha creado el Modelo Genérico de Procesos de Producción de Registros Estadísticos (GSRBPM).

La siguiente figura representa el nuevo modelo GSRBPM desde el punto de vista de la creación/actualización de registros estadísticos que forman parte del sistema integrado de registros estadísticos.

Los procesos y sub-procesos que forman parte del modelo GSRBPM, al igual que el modelo GSBPM, no necesariamente se ejecutan en un orden estricto. Si bien siguen la secuencia lógica representada por la figura 12, ésta podrá variar dependiendo de las circunstancias y algunos de estos procesos o sub-procesos podrían ejecutarse varias veces formando un bucle iterativo.

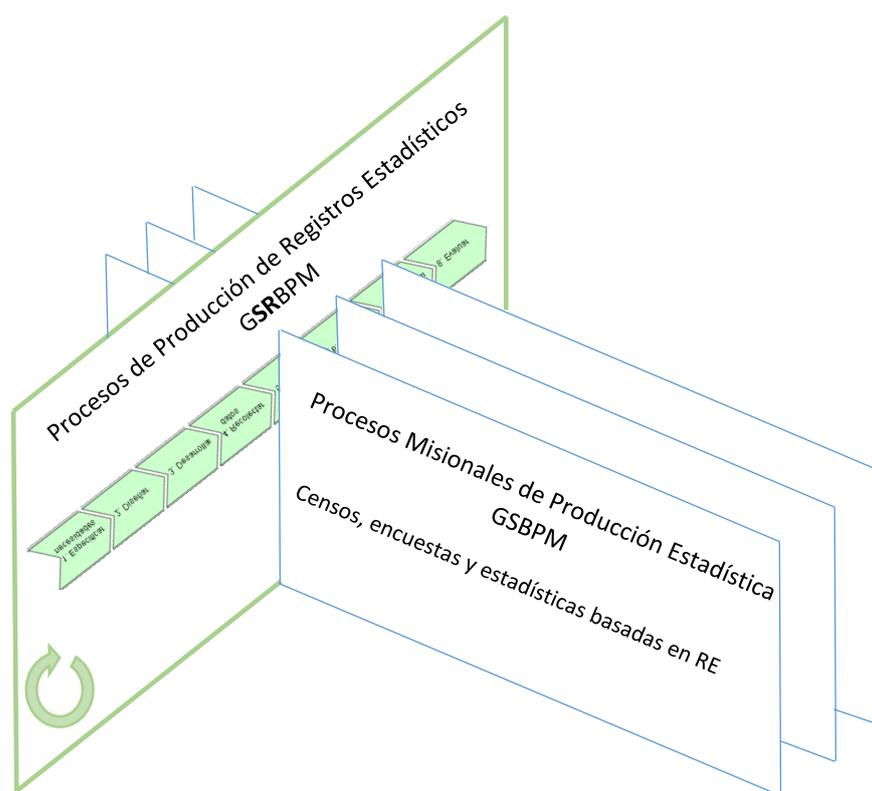
Figura 12. Modelo Genérico de Procesos de Producción de Registros Estadísticos (GSRBPM).



Fuente: elaboración propia del consultor Federico Segui, adaptado del modelo GSBPM de UNECE.

En cambio, si lo vemos desde el punto de vista del sistema de producción de estadísticas en su globalidad lo tendríamos representado de acuerdo a la siguiente figura, donde por un lado tenemos los procesos comunes (base) a todas las áreas temáticas o sectoriales del modelo GSRBPM (Generic Statistical Registers Business Process Model o Modelo Genérico de Procesos de Producción de Registros Estadísticos), (panel verde de la izquierda) que se ejecutan para todos (o la mayoría de) los proyectos estadísticos, y por el otro lado encontramos los procesos llevados a cabo por cada proyecto/investigación (censos, encuestas, estadísticas basadas en registros estadísticos) en una temática específica, de acuerdo al modelo GSBPM (paneles azules de la derecha).

Figura 13. Modelo Genérico Integrado de Procesos de Producción Estadística.



Fuente: elaboración propia del consultor Federico Segui

Descripción de los procesos y sub-procesos que forman parte del modelo **GSRBPM**:

1. Especificar necesidades

Este proceso se lleva a cabo cuando se identifica una necesidad de nuevas estadísticas o cuando la retroalimentación sobre las estadísticas producidas actualmente inicia una revisión. Incluye todas las actividades asociadas con la participación de los usuarios para identificar sus necesidades estadísticas, identificar las posibles fuentes de datos administrativos, crear o actualizar el inventario de registros administrativos y preparar los acuerdos de entrega de información de las fuentes administrativas.

1.1. Identificar necesidades

Este subproceso incluye la investigación inicial y la identificación de las estadísticas que se necesitan y de lo que se necesita de las estadísticas. Puede ser provocado por una nueva solicitud de información, un cambio en el contexto, como una reducción presupuestal o por la necesidad de mejorar los registros base. Los planes de acción a partir de evaluaciones de iteraciones anteriores del proceso, o de otros procesos, podrían proporcionar una entrada a este subproceso.

1.2. Consultar y confirmar necesidades

Este subproceso se centra en la consulta con las partes interesadas y la confirmación en detalle de las necesidades estadísticas. Para que el INE sepa qué se espera producir, cómo y por qué. Para las subsiguientes iteraciones de esta fase, el enfoque principal será determinar si las necesidades previamente identificadas han cambiado. Esta comprensión detallada de las necesidades de los usuarios es la parte crítica de este subproceso.

1.3. Identificar y contactar fuentes administrativas

El objetivo de este subproceso es identificar las posibles fuentes administrativas y los registros administrativos que podrían proveer de los datos necesarios para cubrir las necesidades de los usuarios identificadas y confirmadas en los subprocesos anteriores. Una vez identificadas las fuentes de datos se establece el contacto con la fuente para iniciar el intercambio de información y sobre todo establecer o mejorar el relacionamiento colaborativo mutuo. Se confirma con la fuente de manera primaria y general si dispone en sus registros administrativos de la información requerida para satisfacer las necesidades de los usuarios. Se obtiene información y documentación sobre los registros administrativos de interés.

1.4. Comprobar disponibilidad de datos

Luego de establecer el contacto con la fuente administrativa se comprueba la disponibilidad de los datos necesarios para satisfacer las necesidades de los usuarios y las condiciones en que estarían disponibles, incluidas las restricciones sobre su uso. Se verifica con la fuente si los datos están disponibles en medios electrónicos o deben digitarse, escanearse, etc. Se constata si la fuente está en condiciones de entregar los microdatos identificados del registro administrativo. Cuando se han evaluado las fuentes existentes, se prepara una estrategia para llenar los vacíos restantes en los requerimientos de datos. Este subproceso también incluye una evaluación más general del marco jurídico en el que se recopilarán y utilizarán los datos y, por lo tanto, podrá identificar propuestas de modificación de la legislación vigente o la introducción de un nuevo marco jurídico.

1.5. Evaluar utilidad de la fuente de datos

La utilidad de los datos administrativos es evaluada en términos de cobertura, relevancia, oportunidad y calidad. En el caso que no se considere útil para los propósitos establecidos en el subproceso 1.1, se clasifica la fuente de datos para su uso posterior.

1.6. Crear/actualizar inventario de Registros Administrativos

En este subproceso se crea o actualiza el inventario de registros administrativos y las fuentes responsables. Se registra información básica sobre identificación y denominación de la fuente administrativa y los registros administrativos que gestiona, sus principales variables, cobertura, alcance temático, etc.

1.7. Identificar variables de interés

Se identifican las variables del registro administrativo necesarias para cumplir con los requerimientos estadísticos de los usuarios. Se registran no sólo las variables estadísticas, sino también las variables clave de identificación, variables de referencia temporal, variables de contacto, ubicación geográfica y variables de unión.

1.8. Preparar acuerdo de entrega de información

En este subproceso se elabora el acuerdo de entrega de información con la fuente administrativa. Allí se establece la forma de entrega, formato de los archivos y medio de entrega, periodicidad de las actualizaciones, variables y documentación necesaria.

2. Diseñar

Esta fase está compuesta por las actividades de diseño y desarrollo y cualquier trabajo de investigación práctica necesario para definir los productos, conceptos, metodologías, instrumentos de recolecta y procesos operativos. Incluye todos los elementos de diseño necesarios para definir o refinar los registros estadísticos. Esta fase especifica todos los metadatos relevantes, listos para su uso más adelante en el proceso GSRBPM, así como los procedimientos de aseguramiento de calidad. Esta fase suele ocurrir durante la primera iteración y siempre que se identifican acciones de mejora en la fase Evaluar de un ciclo anterior.

2.1. Definir objetos o unidades estadísticas

Los denominados objetos o unidades estadísticas corresponden a entidades, objetos, elementos o individuos del mundo real, ya sean personas, empresas u organizaciones, hogares, viviendas, inmuebles, vehículos, etc. Son los elementos que forman parte de la población.

Existen diferentes tipos de objetos dentro del sistema de registros estadísticos. Todos los tipos de objetos deben tener una definición clara y precisa, y se deben documentar tanto las definiciones administrativas como las estadísticas.

Se deben documentar las definiciones de las variables clave de identificación de los objetos. Serán utilizadas para detectar duplicados y unir o emparejar registros.

Se debe mantener un registro de referencias cruzadas con todos los cambios de identificación de los objetos a lo largo del tiempo (cambios de número de identificación personal de la población, o cambio del número de predio del inmueble).

2.2. Definir la población

Cuando se genera un nuevo registro estadístico para una investigación específica, se debe definir la población del nuevo registro. Cada registro fuente tiene su propio conjunto de objetos (población), que se incluirá total o parcialmente en el nuevo registro.

Según Wallgren y Wallgren: la definición de una población debe mostrar claramente qué objetos están incluidos en esa población. El tipo de objeto también se especificará con claridad. Asimismo, siempre se incluirán una referencia temporal y una delimitación geográfica. Esta última deberá indicar la relación que existe entre los objetos o unidades estadísticas y el área geográfica.

Se deben usar los registros base para definir los conjuntos de objetos o población del RE por dos razones: los conjuntos de objetos del registro base son los mejores, en teoría, y porque las estadísticas basadas en registros deben ser consistentes.

Requisitos que deben cumplir los registros base para ser utilizados en la definición de poblaciones de los RE:

- Contener referencias temporales (fechas de ocurrencia de los eventos que afectan a los objetos).
- Tener una buena cobertura.
- Tener variables de unión de buena calidad.
- Tener variables de estratificación actualizadas y de buena calidad.

2.3. Definir las variables

Este subproceso define las variables que se utilizarán de los registros administrativos, así como cualquier otra variable que se derivará de ellas en el subproceso 6.3 (derivar nuevas variables y unidades) y cualquier clasificación estadística que se utilizará. Los metadatos de las variables del RA y las derivadas deben documentarse antes de iniciar las fases subsiguientes.

2.4. Diseñar la recolecta de datos

Este subproceso determina el método o métodos de recolecta de datos más apropiados o disponibles con la fuente administrativa. Las actividades reales en este subproceso variarán según el tipo de instrumentos de recolección utilizados, que pueden incluir cuestionarios en papel digitados en el INE, transferencia de archivos de datos, webservices para acceso a los datos administrativos, integración de

bases de datos, etc. Este subproceso incluye el diseño de instrumentos de recolecta. También incluye los ajustes de cualquier acuerdo formal relacionado con el suministro de datos, como memorandos de entendimiento y confirmación de la base legal para la entrega de datos, ambos elementos considerados en el subproceso 1.8 *Preparar acuerdo de entrega de información*, pero que deben ser ratificados o ajustados en esta etapa.

2.5. Diseñar procesamiento e integración

Este subproceso diseña la metodología de procesamiento estadístico que se aplicará durante las fases "Transformar" e "Integrar". Esto puede incluir la especificación de rutinas para codificar, depurar, imputar, integrar, validar y finalizar datasets.

2.6. Diseñar la actualización de registros estadísticos

Este subproceso establece cómo se realiza la actualización de los archivos o tablas de la base de datos de producción estadística correspondientes a los registros estadísticos que se están creando/actualizando (primarios o integrados y base, si corresponde) para su uso por parte de las áreas temáticas o usuarios externos.

2.7. Diseñar sistemas de producción y flujos de trabajo

Este subproceso establece el flujo de trabajo desde la recolecta de datos hasta la comunicación, tomando una visión general de todos los procesos requeridos en todo el proceso de creación y actualización del sistema de registros estadísticos, asegurando que se acoplan eficientemente sin brechas o redundancias. Se necesitan varios sistemas y bases de datos durante todo el proceso. Se debe hacer hincapié en reutilizar procesos y tecnología. Este subproceso también considera cómo el personal interactuará con los sistemas y quién será responsable de qué y cuándo.

3. Desarrollar

En esta fase se construye y prueba la solución hasta el punto en que está lista para su uso en el entorno de producción. Los resultados de la fase "Diseñar" orientan la selección de procesos, instrumentos, información y servicios reutilizables que se montan y configuran en esta fase para crear el entorno operacional completo para ejecutar el proceso. Los nuevos servicios se crean por excepción, y desarrollados en respuesta a las brechas existentes en el catálogo de servicios (sistemas) de la organización y externamente. Estos nuevos servicios serán construidos para ser ampliamente reutilizables dentro de la arquitectura de producción de registros estadísticos.

3.1. Desarrollar o mejorar instrumentos de recolecta de datos

Este subproceso abarca las actividades para desarrollar los instrumentos de recolecta que se utilizarán durante la fase "Recolectar datos". El

instrumento de recolecta se genera o construye sobre la base de las especificaciones de diseño creadas durante la fase "Diseñar". Este subproceso también incluye preparar y probar el funcionamiento de ese instrumento. Se recomienda asociar los instrumentos de recolecta al sistema de metadatos estadísticos, de modo que los metadatos puedan capturarse más fácilmente en la fase "Recolectar datos". La conexión de metadatos y datos en el momento de la recolecta puede ahorrar trabajo en fases posteriores.

3.2. Desarrollar o mejorar componentes del procesamiento

En este subproceso llevan a cabo las actividades para desarrollar nuevos y mejorar componentes y servicios existentes necesarios para la fase "Transformar", tal como se diseñó en la fase "Diseñar". Esto incluye el desarrollo de rutinas para estandarizar variables, detectar duplicados, codificar, depurar, imputar y validar datasets.

3.3. Desarrollar o mejorar componentes de integración

Este subproceso incluye las actividades para desarrollar nuevos y mejorar componentes y servicios existentes necesarios para la fase "Integrar", tal como se diseñó en la fase "Diseñar". Esto incluye el desarrollo de rutinas para integrar RA y RE, derivar nuevas variables y unidades, des-identificar registros (proteger datos), actualizar y finalizar datasets.

3.4. Configurar flujos de trabajo

Este subproceso configura el flujo de trabajo, los sistemas y las transformaciones utilizados en los procesos de creación y actualización del sistema de registros estadísticos, desde la recolecta de datos hasta la comunicación. Asegura que el flujo de trabajo especificado en el subproceso 2.6 *Diseñar sistemas de producción y flujo de trabajo* funcione en la práctica.

3.5. Probar sistemas de producción

Este subproceso se refiere a la prueba de los sistemas implementados y configurados y sus respectivos flujos de trabajo. Incluye las pruebas técnicas y la aprobación de nuevos programas y rutinas, así como la confirmación de que las rutinas existentes de otros procesos de negocio estadísticos son adecuadas para su uso en este caso. Mover los componentes del proceso al entorno de producción y asegurar que funcionen como se espera en ese entorno

3.6. Probar procesos (pruebas piloto)

Este subproceso abarca las actividades para gestionar una prueba piloto del proceso de actualización y mantenimiento del sistema de registros estadísticos. Normalmente incluye una recolecta de datos a pequeña escala, para probar los instrumentos de recolecta, seguido por el

procesamiento e integración de los datos, para asegurar que el proceso se ejecute como se esperaba. Después del piloto, puede ser necesario volver a pasos anteriores y hacer ajustes a instrumentos, sistemas o componentes.

3.7. Finalizar sistemas de producción

Este subproceso incluye las actividades para que los procesos y sistemas implementados y configurados, incluyendo los servicios modificados y recién creados en producción, queden listos para el uso por las áreas temáticas. Las actividades incluyen: producir documentación sobre los componentes del proceso, incluida la documentación técnica y los manuales de usuario; y capacitar a los usuarios sobre cómo operar el proceso.

4. Recolectar datos

Esta fase recoge o reúne toda la información necesaria (datos y metadatos), utilizando diferentes modos de recolección (digitación de formatos papel, extracción de datos de archivos, integración de bases de datos, webservices, etc.) y los carga en el entorno adecuado para su posterior procesamiento. Aunque puede incluir la validación de formatos del archivo de datos (dataset), no incluye ninguna transformación de los datos en sí, ya que esto se realiza en la fase "Transformar".

4.1. Organizar/preparar recolecta de datos

Este subproceso garantiza que las personas, los procesos y la tecnología están preparados para recopilar datos y metadatos. Este subproceso incluye:

- Garantizar la disponibilidad de recursos de recolecta, por ejemplo: servidores, medios de transferencia de archivos, software de extracción de datos, sistemas de integración de bases de datos.
- Configurar sistemas de recolecta para solicitar y recibir los datos.
- Asegurar la confidencialidad de los datos a recolectar.
- Preparar instrumentos de recolecta.

4.2. Ejecutar recolecta de datos

Este subproceso implica la recolecta de datos en sí misma. Se ponen en marcha las actividades de recolecta de datos utilizando los instrumentos especialmente diseñados y desarrollados para recolectar los datos provistos por las fuentes administrativas. Se realiza la extracción de datos (lectura de archivos, integración de BD, webservices, etc.).

4.3. Finalizar recolecta de datos

Este subproceso incluye la carga de los datos y metadatos recopilados en un entorno electrónico adecuado (base de datos de entrada, con datos crudos) para su posterior procesamiento. Puede incluir la toma de datos

manuales o automáticos, por ejemplo, utilizando personal administrativo u herramientas ópticas de reconocimiento de caracteres para extraer información de cuestionarios en papel o convertir los formatos de archivos recibidos de las fuentes administrativas. También puede incluir análisis de los metadatos del proceso (paradata) asociados con la recolecta para asegurar que las actividades de recolección cumplan con los requisitos.

4.4. Evaluar calidad de registros administrativos

Se debe evaluar la calidad de los datos de los registros administrativos una vez que son recolectados. Se debe seguir la metodología de evaluación de la calidad de los datos del registro y se debe generar el correspondiente reporte de la calidad. Este subproceso será clave para determinar efectivamente la utilidad de los datos y sus restricciones en cuanto al uso estadístico.

5. Transformar

Esta fase implica la transformación del registro administrativo en registro estadístico y su preparación para la integración con otros registros. Se compone de subprocesos que estandarizan variables, validan, depuran y transforman los datos de entrada, para que puedan ser utilizados con fines estadísticos. Puede repetirse varias veces si es necesario. Para las actualizaciones de datos recolectadas regularmente, esta fase se produce en cada iteración.

5.1. Estandarizar/armonizar variables

Las definiciones de las variables del RA se ajustan a las necesidades administrativas y no siempre se corresponden con las definiciones estadísticas de acuerdo a un uso estadístico específico.

La estandarización o armonización de variables implica la adecuación de las definiciones de las variables del RA a las definiciones estandarizadas del RE. Para lo cual podría ser necesario transformar las categorías de respuesta de variables, los clasificadores/codificadores utilizados, los formatos de datos o los datos en sí mismos. Además, implica la normalización de textos y creación o actualización de fichas de metadatos.

5.2. Revisar y validar

Este subproceso examina los datos para tratar de identificar problemas potenciales, errores y discrepancias tales como valores atípicos, falta de respuesta en las variables y errores de codificación. También puede denominarse validación de datos de entrada. Puede ejecutarse iterativamente, validando datos contra reglas de inconsistencias predefinidas, generalmente en un orden establecido. Puede marcar los datos para la revisión o edición automática o manual. La revisión y

validación puede aplicarse a los datos de cualquier tipo de fuente, antes y después de la integración de datos. Si bien la validación se trata como parte de la fase "Transformar", en la práctica, algunos elementos de validación pueden ocurrir junto con las actividades de recolecta y luego de la fase "Integrar", cuando se tienen los registros unidos y se pueden utilizar variables adicionales para las validaciones. Si bien este subproceso se refiere a la detección de errores reales o potenciales, cualquier actividad de corrección que realmente cambie los datos se realiza en el subproceso 5.3.

5.3. Depurar e imputar

Cuando los datos se consideran incorrectos, faltan o no son fiables, este subproceso puede incorporar nuevos valores en su lugar. Los términos depuración/edición e imputación cubren una variedad de métodos para hacer esto, a menudo usando un enfoque basado en reglas. Los pasos específicos típicamente incluyen:

- Determinación de agregar o cambiar datos.
- Selección del método a utilizar.
- Agregar / cambiar valores de datos.
- Escribir los nuevos valores de los datos en el conjunto de datos y marcarlos como cambiados.
- Documentación de metadatos sobre el proceso de depuración/edición e imputación.

5.4. Clasificar/codificar variables

Este subproceso clasifica y codifica los datos de entrada. Por ejemplo, las rutinas de codificación automáticas (o manuales) asignan códigos numéricos a las respuestas de preguntas de texto abiertas de acuerdo con un esquema de clasificación predeterminado.

5.5. Detectar duplicados

Los archivos del RA pueden contener duplicados (casos o filas con los mismos datos, claves de identificación duplicada, claves diferentes pero los casos corresponden al mismo objeto o unidad). Se utilizan métodos determinísticos y probabilísticos para detectar duplicados.

Se detectan duplicados a través de las claves de identificación, que luego serán eliminados (borrado lógico, se marca la fila con el estado *inactivo por duplicado*) si a través de los métodos probabilísticos se detecta que hay una alta probabilidad de que se trate del mismo objeto o unidad en la realidad.

Las actividades previas a la aplicación de los métodos probabilísticos son: normalización de textos, transformación de variables numéricas o fecha a texto, concatenación de variables (también se utilizan las variables clave de identificación).

Luego, se aplican las técnicas probabilísticas de unión de registros (el archivo del RA se empareja con sí mismo para detectar duplicados), donde a cada caso/fila del archivo se le asigna una probabilidad de coincidencia (posible duplicado) con un caso/fila del mismo archivo. Se aplican técnicas de blocking, indexing o filtering, para trabajar con bloques del archivo de menor tamaño y mejorar la velocidad de procesamiento.

Las pares de casos cuya probabilidad de coincidencia supera cierto umbral son considerados como duplicados y serán marcados como eliminados (borrado lógico, se marca la fila con el estado *inactivo por duplicado*) del archivo de datos.

Finalmente, la calidad de los resultados es evaluada.

6. Integrar

En esta fase se realiza la integración de registros a través de métodos de unión o emparejamiento de archivos. Se crean nuevas variables y unidades luego de la unión de registros. Se des-identifican los registros para asegurar la confidencialidad de la información previamente a la actualización en la base de datos de producción estadística.

6.1. Integrar registros (unir o emparejar)

Se aplican métodos determinísticos y probabilísticos para realizar la unión de registros.

La unión de registros consta de tres etapas: pre-unión, unión y post-unión. La pre-unión corresponde a la normalización de variables de texto y estandarización otras variables (realizada en la fase 5 "Transformar"). La unión de registros se realiza aplicando algoritmos estadísticos estandarizados y ampliamente utilizados en diversas disciplinas. Finalmente, la post-unión o proceso de evaluación de los resultados (siguiente subproceso).

6.2. Evaluar resultados de la integración

La evaluación de resultados de la integración permite identificar oportunidades de mejora de los procesos y métodos utilizados, para lograr mejores tasas de coincidencias, minimizando los "falsos positivos" incluidos en las coincidencias y los "falsos negativos" excluidos del emparejamiento.

6.3. Derivar nuevas variables y unidades

Este subproceso deriva los datos de variables y unidades que no se proporcionan explícitamente en el archivo del registro administrativo, pero son necesarios para cumplir con los requerimientos de los usuarios. Se crean nueva variables a partir de otras variables del RA o de los RE

con los que se ha integrado (se crean por varios métodos diferentes). Esta actividad podría ser iterativa, ya que algunas variables derivadas pueden basarse, a su vez, en otras variables derivadas. Por lo tanto, es importante asegurarse de que las variables se derivan en el orden correcto. Las nuevas unidades pueden derivarse agregando o dividiendo datos de las unidades del RA, o por otros métodos de estimación. Los ejemplos incluyen los hogares derivados de las unidades del RA de personas o las empresas donde las unidades del RA son unidades legales.

6.4. Aplicar la protección de datos

En este subproceso se hace la desidentificación del registro estadístico. Es decir, se sustituyen las variables clave de identificación por claves aleatorias y se eliminan las variables de contacto o identificadores explícitos (nombres, teléfono, dirección, email, coordenadas geográficas) del RE que estará disponible para los usuarios internos y externos.

ATENCIÓN: este proceso de protección de datos simplemente hace la des-identificación de registros. No se trata de un proceso de anonimización completo, el cual debería implementarse antes de liberar una base de microdatos a usuarios externos.

6.5. Crear/actualizar registro estadístico en BD de producción estadística

Este subproceso realiza la actualización de los archivos o tablas de la base de datos de producción estadística correspondientes a los registros estadísticos que se están creando/actualizando (primarios o integrados y base, si corresponde), para su uso por parte de las áreas temáticas o usuarios externos. Se actualiza el estado de los casos/filas del registro (Activo, Inactivo por emigración, Inactivo por muerte, Eliminado por proceso administrativo, Inactivo por duplicado, etc.).

6.6. Crear/actualizar registro base

Se evalúa la necesidad de actualizar o mejorar el registro base. Se seleccionan las variables y unidades a actualizar o sustituir en el registro base para mejorar su cobertura, calidad y actualidad de los datos. Se actualiza la ficha de metadatos del registro base, y se verifica que se actualiza la información correspondiente a la trazabilidad o mapeo de variables.

Se actualiza el estado de los casos/filas del registro base (Activo, Inactivo por emigración, Inactivo por muerte, Eliminado por proceso administrativo, Inactivo por duplicado, etc.).

7. Comunicar

Es importante comunicar los cambios y actualizaciones de datos realizadas en el sistema de RE y en los sistemas de difusión y exploración de datos, de

modo tal que los usuarios internos y externos conozcan el nuevo alcance y posibilidades del sistema en cuanto a la producción de estadísticas.

7.1. Actualizar sistemas de difusión de datos y herramientas de análisis

Este subproceso gestiona la actualización de los sistemas en los que los datos y los metadatos se almacenan listos para su difusión o explotación, entre ellos:

- Dar formato a los datos y metadatos para que estén listos para ser cargados en bases de datos de salida;
- Cargar datos y metadatos en las bases de datos de salida;
- Asegurar que los datos estén vinculados a los metadatos relevantes.
- Actualizar el Data Warehouse (si corresponde).

El formateo, la carga y la vinculación de los metadatos deberían realizarse principalmente en fases anteriores, pero este subproceso incluye una verificación final de que todos los metadatos necesarios están listos para su difusión o explotación.

7.2. Gestionar atención al usuario

Este subproceso garantiza que se registran las consultas de los usuarios y las peticiones de servicios como el acceso a los microdatos, y que las respuestas se proporcionen dentro de los plazos acordados. Estas consultas y solicitudes deben revisarse periódicamente para proporcionar una entrada al proceso de gestión de calidad general, ya que pueden indicar cambios o nuevas necesidades de los usuarios.

7.3. Comunicar actualizaciones del sistema de registros estadísticos

Se deben comunicar adecuadamente los cambios y actualizaciones realizadas en los datos del sistema de RE, de modo tal que los usuarios internos y externos conozcan el nuevo alcance y posibilidades del sistema en cuanto a la producción de estadísticas

8. Evaluar

Esta fase gestiona la evaluación de un caso o ciclo específico del proceso de gestión del sistema de RE, en contraposición con el proceso general de gestión de la calidad que abarca a todos los procesos del INE. Lógicamente tiene lugar al final de la instancia del proceso, pero se basa en los insumos reunidos a lo largo de las diferentes fases. Incluye la evaluación del desempeño de una instancia específica del proceso, considerando una serie de insumos cuantitativos y cualitativos, y la identificación y priorización de mejoras potenciales.

Para los procesos de actualización regular de los datos de los mismos RA, donde se ejecutan ciclos iterativos, la evaluación debería (al menos en teoría)

ocurrir para cada iteración y si se deberían implementar mejoras en los siguientes ciclos.

8.1. Reunir información para la evaluación

El material para la evaluación se puede producir en cualquier otra fase o subproceso. Puede tomar muchas formas, incluyendo retroalimentación de los usuarios, metadatos del proceso (paradata), métricas del sistema y sugerencias del personal. Los informes de avances con respecto a un plan de acción acordado durante una iteración anterior también pueden ser un aporte a las evaluaciones de las iteraciones posteriores. Este subproceso reúne todos estos insumos y los pone a disposición de la persona o equipo que realiza la evaluación.

8.2. Conducir evaluación

Este subproceso analiza los insumos de la evaluación y los sintetiza en un informe de evaluación. El informe resultante debe incluir cualquier problema de calidad específico de esta iteración del proceso y debe hacer recomendaciones de mejora, si corresponde. Estas recomendaciones podrían considerar cambios en cualquier fase o subproceso para futuras iteraciones del proceso, o podrían sugerir que el proceso no se repita.

8.3. Elaborar y acordar plan de mejora

Este subproceso reúne a quienes tienen el poder de decisión necesario para elaborar y acordar un plan de acción basado en el informe de evaluación. También debería incluir la consideración de un mecanismo para monitorear el impacto de esas acciones, las cuales, a su vez, pueden aportar una contribución a las evaluaciones de futuras iteraciones del proceso.

8.4. Implementar plan de mejora

Las acciones acordadas en el plan de mejora se implementan de acuerdo a las prioridades y cronograma establecidos. Los responsables designados en el plan coordinan la correcta ejecución de las acciones de mejora.

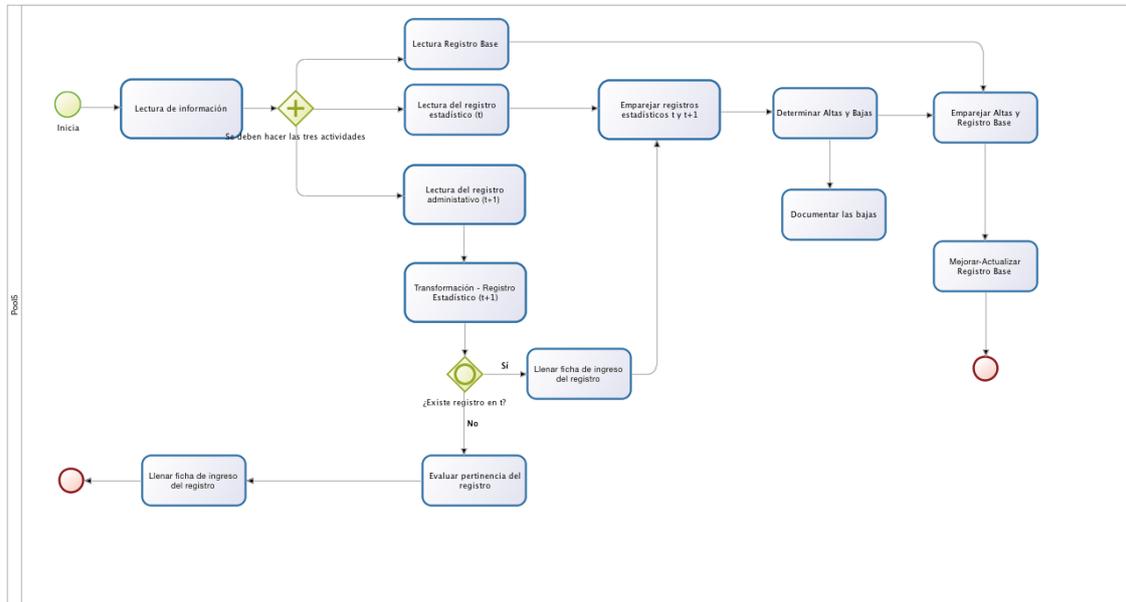
6.3. Diseño de procesos

El diseño y descripción de procesos se ha realizado siguiendo el estándar internacional de notación gráfica que permite el modelado de proceso de negocio en un formato de flujo de trabajo: BPMN (Business Process Model and Notation).

Esta herramienta es utilizada por los países de la Comunidad Andina y en particular el DANE de Colombia y el INEC de Ecuador han descrito gráficamente los procesos de transformación, actualización e integración de

registros estadísticos aplicando esta notación. Por ejemplo, el DANE ha diseñado un flujo de trabajo correspondiente al proceso de actualización del RE, el cual está representado por la siguiente figura.

Figura 14. Proceso de actualización de registros



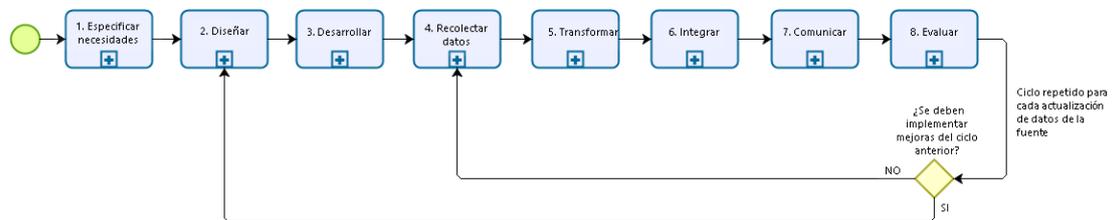
Fuente: DANE de Colombia.

A continuación se describe el flujo del proceso de producción estadística basado en un sistema integrado de registros estadísticos, descrito conceptualmente por el modelo GSRBPM.

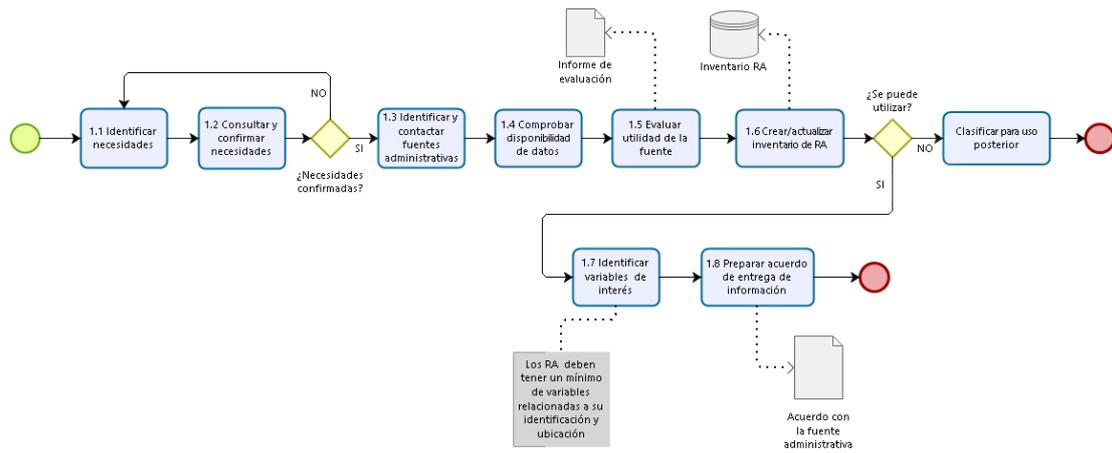
La siguiente figura corresponde a los diagramas de procesos y subprocesos desarrollados con la herramienta de BPMN para representar gráficamente el proceso GSRBPM.

Figura 15. Diagrama de procesos del modelo GSRBPM según la notación BPMN.

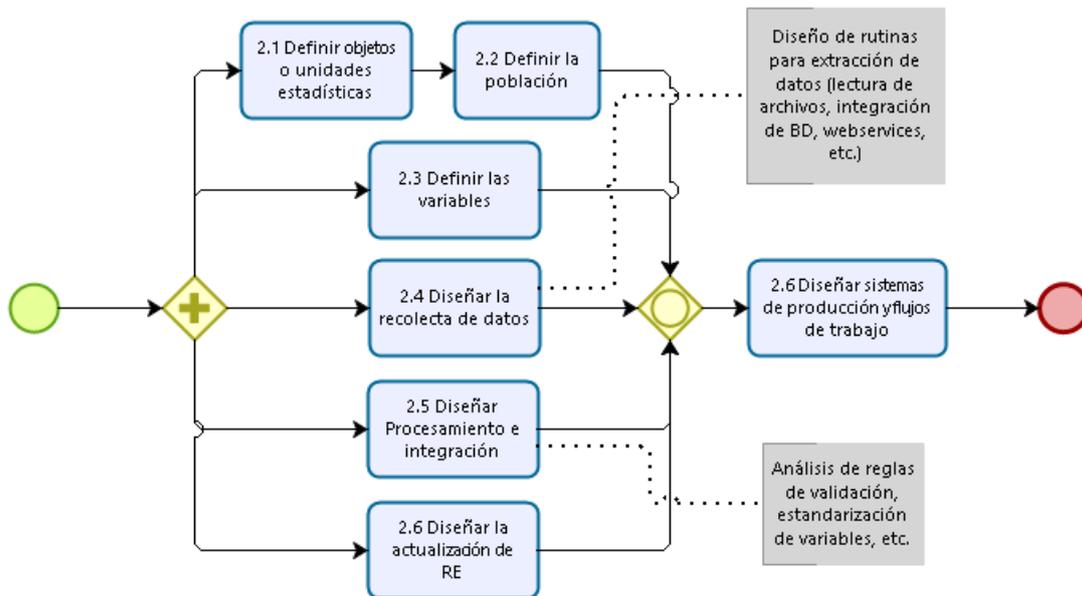
Diagrama de primer nivel – Fases del modelo GSRBPM



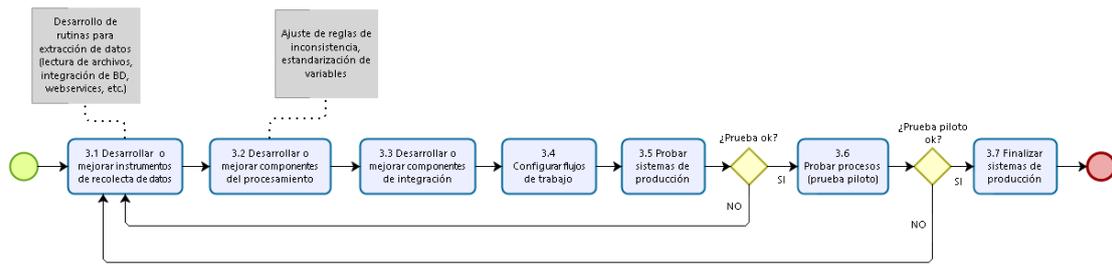
1. Especificar necesidades



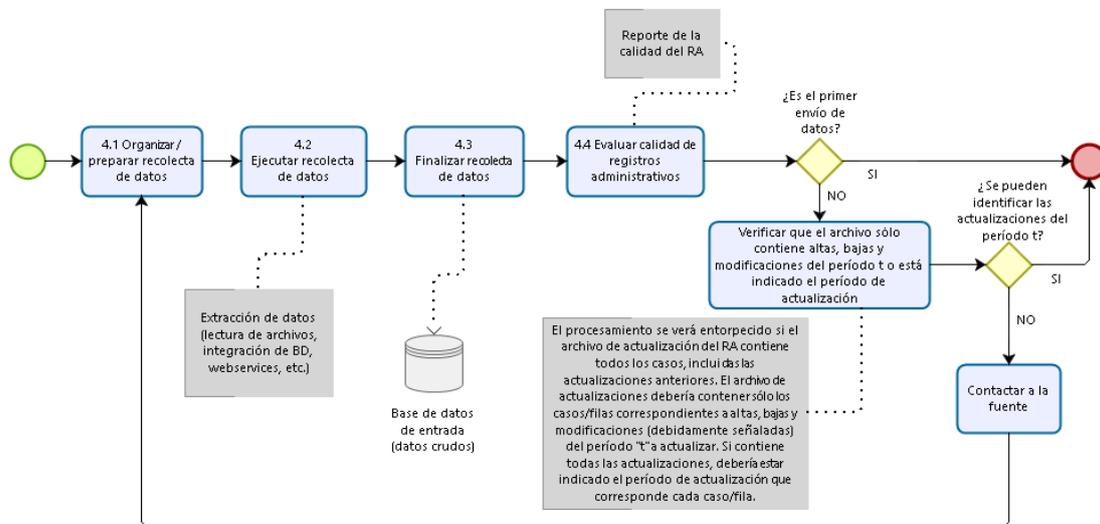
2. Diseñar



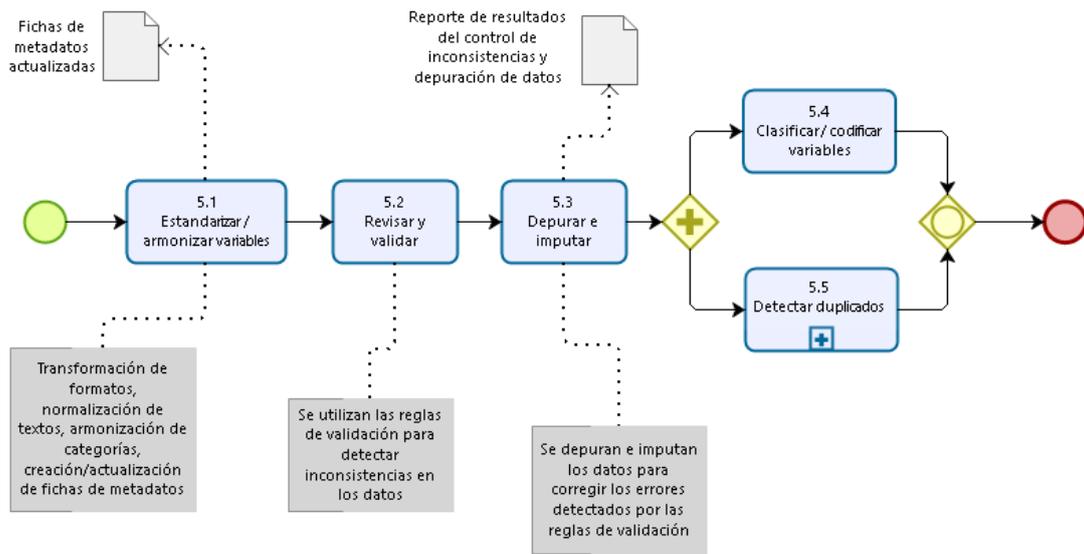
3. Desarrollar



4. Recolectar datos

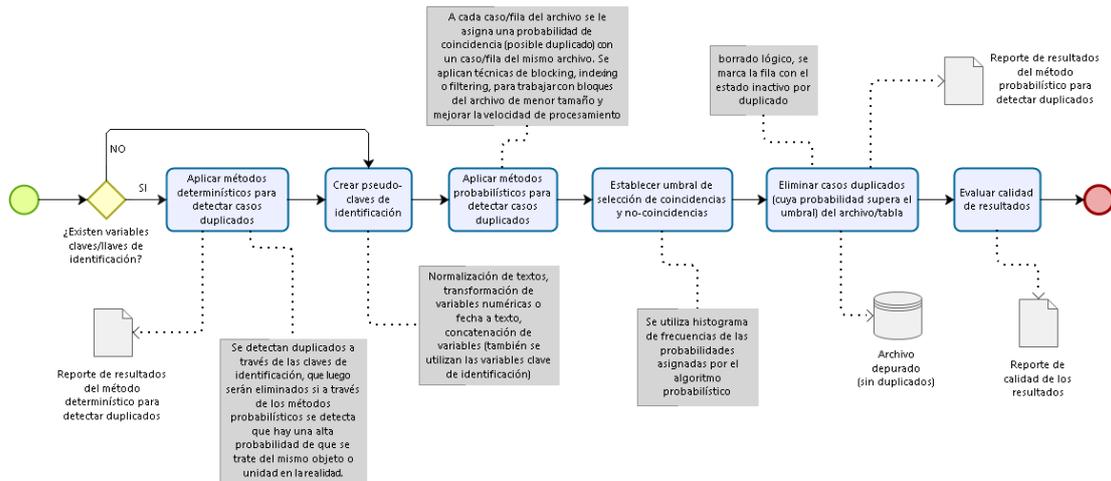


5. Transformar



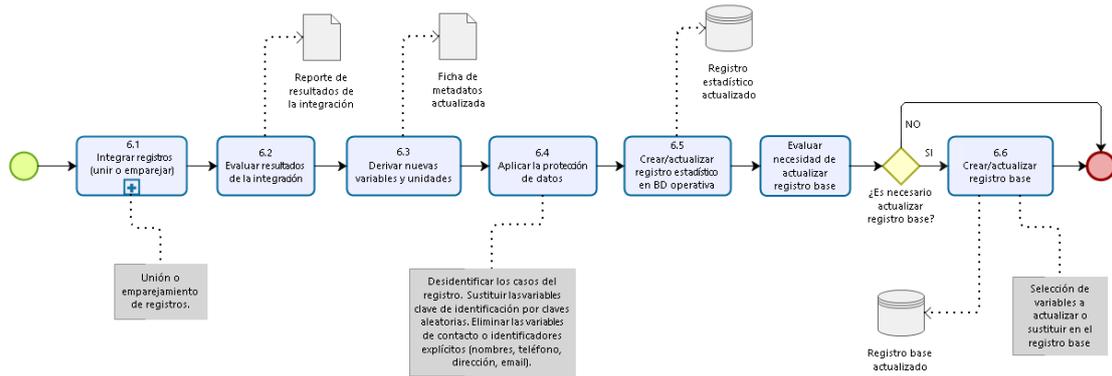
Powered by
bizagi
 Modeler

5.1 Estandarizar/armonizar variables

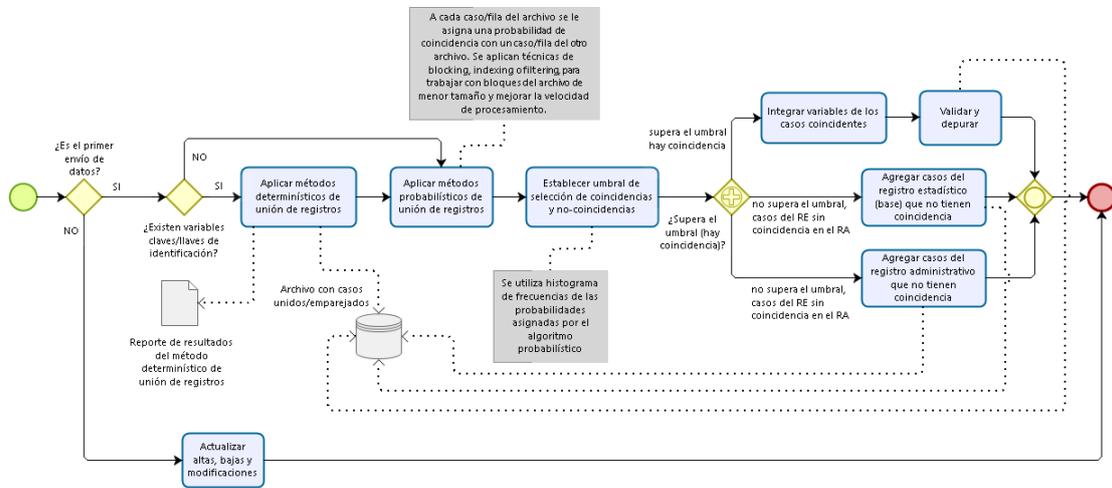


Powered by
bizagi
 Modeler

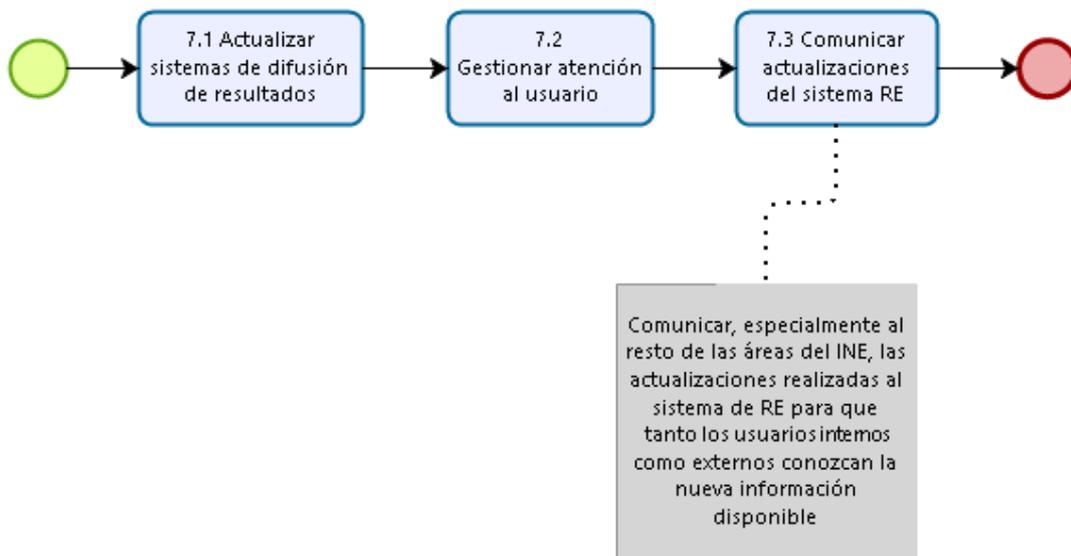
6. Integrar



6.1 Integrar RA y registros estadísticos

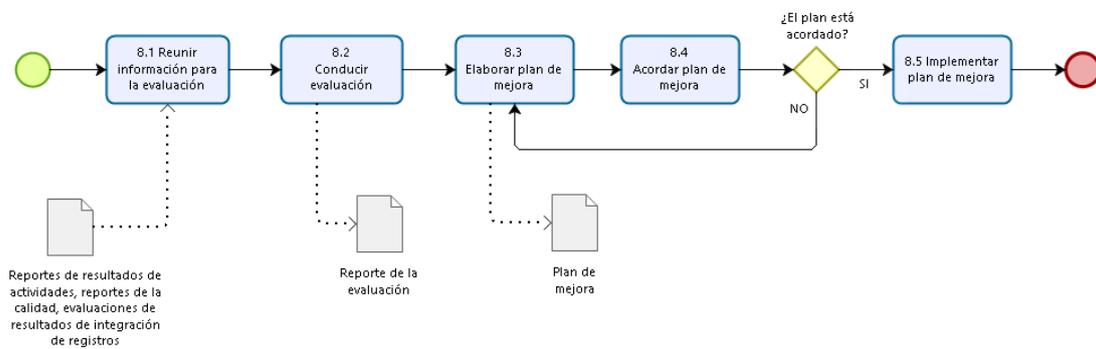


7. Comunicar



Powered by
bizagi
Modeler

8. Evaluar



Powered by
bizagi
Modeler

7. Aspectos organizacionales

La implementación de un modelo de producción de estadísticas basadas en un sistema integrado de registros estadísticos implica cambios importantes tanto en los procesos como en la estructura organizativa de la institución.

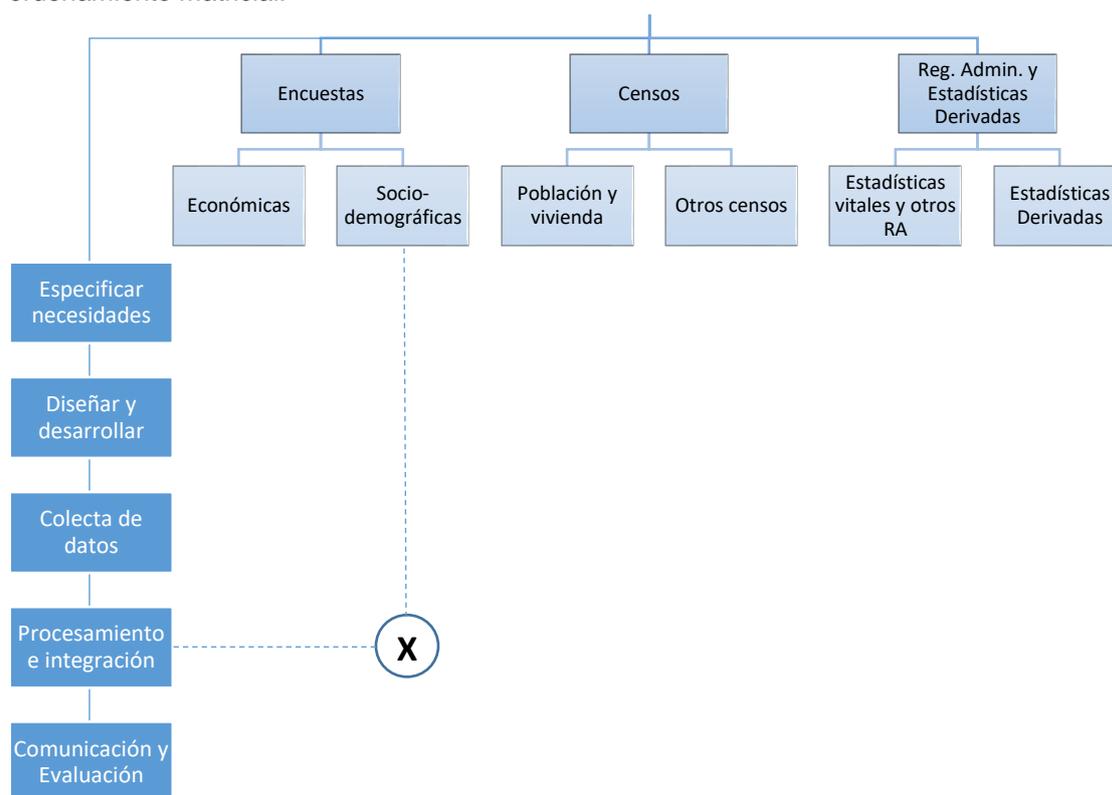
La adecuada gestión del cambio incidirá en el éxito del proyecto. La comunicación de los cambios y del plan de acción dentro de la organización será fundamental para que la transición se logre de manera armoniosa y con el menor impacto posible, tratando de minimizar la natural resistencia al cambio que ofrecen las personas ante nuevos retos y desafíos.

La gestión del sistema integrado de registros estadísticos supone la asignación de recursos humanos dedicados exclusivamente a esta actividad. En este sentido, será necesario crear un área dentro de la estructura organizativa o un grupo de trabajo abocado a la implementación, coordinación, administración y mantenimiento del sistema.

En este punto se ahondará en los aspectos organizacionales relativos a la gestión del sistema integrado de registros estadísticos.

La siguiente figura representa la estructura organizativa que deberían adoptar los INE, donde las áreas temáticas y los procesos se combinan de forma matricial (Medina, Seguí (2013). *¿Cómo mejorar el desempeño y crear valor público en las oficinas y sistemas estadísticos nacionales en América Latina y el Caribe?*), para lograr así una eficaz y eficiente ejecución de procesos y actividades.

Figura 16. Estructura organizativa del proceso productivo de registros estadísticos con ordenamiento matricial.



Este enfoque matricial combina la estructura de “departamentalización” temática o sectorial que generalmente tienen las oficinas de estadística con el enfoque de gestión por procesos determinado por el modelo **GSRBPM** (ilustrado en la figura 12).

En este modelo organizacional existen áreas encargadas de ejecutar procesos comunes a todas las áreas temáticas o sectoriales (en ciertos puntos también son procesos transversales que involucran a todas estas áreas), como por ejemplo el procesamiento e integración de registros administrativos y registros estadísticos.

Según este ejemplo, los registros administrativos son procesados (estandarización, codificación, depuración de variables) para transformarlos en registros estadísticos e integrarlos con los registros estadísticos ya existentes en el INE. Este proceso no es ejecutado exclusivamente para un proyecto de un tema o sector en particular, sino que su propósito es actualizar el sistema integrado de registros estadísticos en forma continua. Sin embargo, se producen interacciones con las áreas temáticas y sus procesos productivos, pues obviamente sus necesidades son tenidas en cuenta a la hora de integrar las fuentes de información. Estos puntos de contacto están representados por la “X” en la figura anterior (figura 16).

El modelo **GSRBPM** rompe con los silos que generalmente se forman en los INE en torno a las áreas temáticas o sectoriales. En este enfoque las áreas temáticas

se ven obligadas a interactuar con las áreas encargadas de los procesos comunes de GSRBPM. Es así que los procesos GSBPM de cada proyecto o investigación estadística tienen puntos de contacto con los procesos del GSRBPM.

La propuesta es crear un área encargada de la ejecución y seguimiento de los procesos (comunes) del sistema integrado de registros estadísticos (según el modelo GSRBPM) y de los mecanismos de coordinación de los puntos de contacto con los procesos del GSBPM de cada proyecto de los diferentes sectores.

La coordinación y comunicación entre las áreas es un elemento crítico en un modelo organizacional del tipo matricial. Si falla la coordinación entre las áreas temáticas o sectoriales y el área encargada de los procesos del sistema integrado de registros estadísticos fallará el sistema como tal.

El área encargada del sistema integrado de registros estadísticos (según modelo GSRBPM) debería liderar grupos de trabajo integrados por representantes de las áreas temáticas o sectoriales. Los líderes de cada grupo deberían ser especialistas sectoriales con capacidad de coordinación y liderazgo de grupos de trabajo.

8. Marco conceptual y metodológico común

En este capítulo se desarrolla el marco metodológico común que sustenta el diseño, desarrollo e implementación de un sistema integrado de registros estadísticos de población e inmuebles en la subregión Andina.

El marco conceptual y metodológico está basado en la metodología de los países nórdicos, tal como se ha mencionado en capítulos anteriores. Esta metodología establece un sistema de registros conformado por cuatro registros base, pero en el caso del proyecto BPR serán considerados solo los registros de población e inmuebles.

Además, han sido consideradas las experiencias y avances realizados por los países de la Comunidad Andina como base de la propuesta.

La factibilidad de implementar un sistema integrado de registros estadísticos no es la misma para los cuatro países de la subregión y depende de las siguientes precondiciones⁹:

- La existencia de fuentes administrativas adecuadas.
- Facilidad de acceso a los registros (microdatos) administrativos identificados.

⁹ UNECE (2011). Using administrative and secondary sources for official statistics: A handbook of principles and practices. Naciones Unidas. Nueva York y Ginebra, 2011.

- Identificadores comunes.
- Aceptación pública.

Este marco conceptual y metodológico debe ser revisado en forma continua para detectar oportunidades de mejora.

8.1. Elementos del sistema

Este apartado contiene una descripción de los elementos mínimos que forman parte del sistema integrado de registros estadísticos.

Los denominados **objetos**, **unidades** o **elementos** del sistema de registros corresponden a entidades, objetos, elementos o individuos del mundo real, ya sean personas, empresas u organizaciones, hogares, viviendas, inmuebles, vehículos, etc. Son los elementos que forman parte de la población. Por ejemplo, en una población de individuos cuyo tipo de objeto es *persona*, cada individuo de la población es una instancia de objeto (representado por casos o filas en un sistema de archivos o base de datos).

Se debe mantener un registro de referencias cruzadas con todos los cambios de identificación de los objetos a lo largo del tiempo (cambios de número de identificación personal de la población, o cambio del número de predio del inmueble).

La **población** se define por el conjunto de objetos o unidades que la componen, incluyendo información sobre lugar y tiempo. Una definición clara del tipo de objeto (qué significa hogar, empresa, inmueble, etc.) redundará en una definición clara de la población.

La **población de interés o población objetivo** son todos los casos o unidades que forman parte del Registro Estadístico que cumplen con un conjunto de características particulares, es decir, ciertas variables tienen determinados valores en común. La población de interés puede o no coincidir con la totalidad de unidades que componen el Registro Estadístico. Es posible, entonces, definir a la población de interés como un subconjunto del Registro Estadístico, esto va a depender del uso con fines estadísticos que se quiera hacer del mismo.

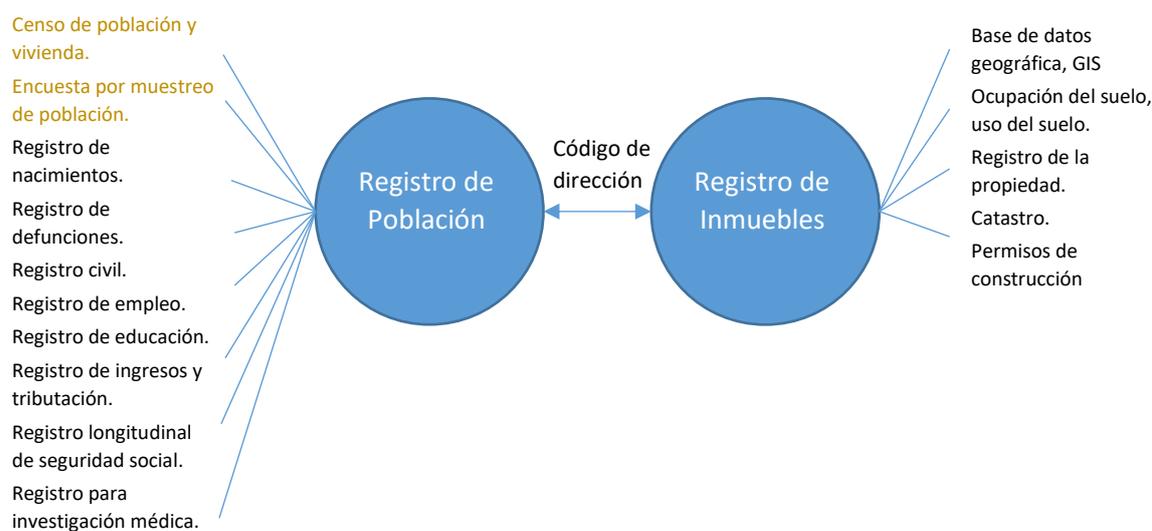
Un **registro estadístico** es el registro consolidado de datos estandarizados y procesados provenientes de uno o más Registros Administrativos, que originalmente no (necesariamente) fueron captados con fines estadísticos, pertenecientes a una o más fuentes de datos administrativos, pero que han sido adaptados para su uso estadístico.

En el siguiente apartado se describen las diferentes categorías de registros estadísticos que componen el sistema.

Los **casos** o **filas** de los registros (administrativos o estadísticos) se refieren a cada uno de los objetos o unidades individuales (instancias de objetos), contienen información individualizada de cada elemento que compone el registro.

La figura 17 muestra en forma resumida los elementos del sistema de registros estadísticos por tipo de objeto y sus fuentes de información, acotado a los registros de población e inmuebles según se ha determinado para el proyecto BPR.

Figura 17. Sistema de registros estadísticos por tipo de objeto y sus fuentes de información administrativa y estadística, acotado a registros de población e inmuebles.



Fuente: Adaptado de Anders Wallgren, Britt Wallgren. (2012). *Estadísticas basadas en registros. Aprovechamiento estadístico de los registros administrativos*. INEGI.

Los objetos o unidades estadísticas tienen una serie de atributos medibles denominados **variables**. A nivel de los casos o filas correspondientes a las instancias de los objetos o unidades las variables son representadas mediante **campos**. Estos campos están organizados en columnas en la matriz de datos que representa el archivo del registro.

*Una **variable estadística** está definida por el tipo de objeto que presenta la característica (por ejemplo, ingreso para personas e ingreso para hogares son dos variables distintas), por el método de medición y la escala aplicados, y por el momento o período a los que refiere la medición.*

En el apartado 8.3.2 se hace una descripción detallada de las diferentes categorías de variables que componen el sistema de registros estadísticos.

8.2. Categorías e interrelaciones de Registros Estadísticos

El sistema integrado de registros estadísticos está conformado por una serie de registros estadísticos: *registros base*, *registros primarios* y *registros integrados*.

Recordemos que un **registro estadístico** es el registro consolidado de datos estandarizados y procesados provenientes de uno o más Registros Administrativos, que originalmente no (necesariamente) fueron captados con fines estadísticos, pertenecientes a una o más fuentes de datos administrativos, pero que han sido adaptados para su uso estadístico

Wallgren y Wallgren señalan que los **registros base** tienen como función definir los objetos y poblaciones del sistema de registros. Son la columna vertebral del sistema pues contienen los tipos de objetos y los vínculos más relevantes. La calidad del sistema está determinada por las definiciones de los objetos y la cobertura de los registros base.

Un registro base tiene las siguientes características (Wallgren y Wallgren):

- *“Define tipos de objetos importantes.*
- *Define conjuntos de objetos o poblaciones estandarizadas importantes.*
- *Contiene vínculos con objetos de otros registros base.*
- *Contiene vínculos con otros registros relacionados con el mismo tipo de objeto.*
- *Es importante para el sistema en su conjunto, por lo que resulta esencial que sea de alta calidad y esté bien documentado.*
- *Es importante para el marco muestral.*
- *Se puede usar para estadísticas demográficas relacionadas con personas, actividades, inmuebles o empresas.*

Las fechas de nacimiento y defunción deben estar presentes en el registro base para producir estadísticas demográficas.”

En el marco del proyecto BPR el sistema integrado de registros estadísticos estará conformado (inicialmente) por dos registros base: población e inmuebles.

Los **registros primarios** son aquellos que se basan directamente en al menos una fuente administrativa.

Los **registros integrados** son creados combinando exclusivamente información ya existente en los registros estadísticos del sistema. Los *registros longitudinales* son un caso particular de registros integrados, los cuales consolidan información

de varios registros estadísticos anuales para hacer seguimiento a los mismos objetos o unidades a lo largo del tiempo.

Otro tipo de registros administrativos utilizados en determinadas ocasiones son los denominados **registros satélite** o **asociados**. Son registros que están disponibles para los INE, contienen información acerca de unidades y variables de interés, y tienen las siguientes características:

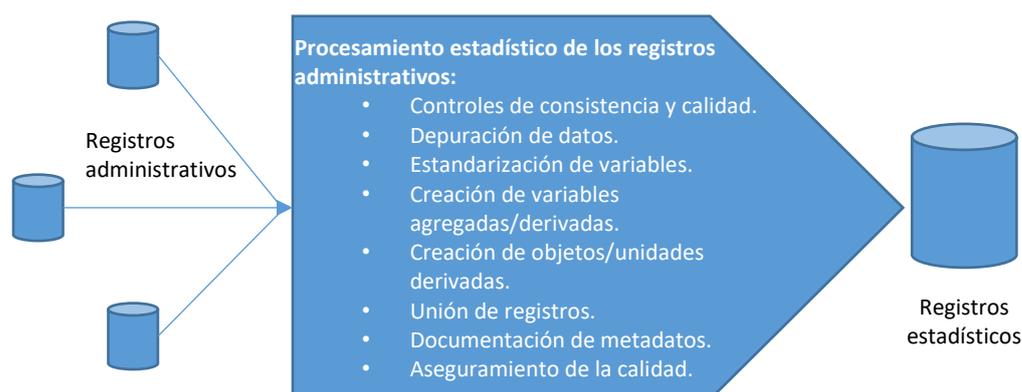
- No son parte integral del sistema de registros, pero son capaces de ser vinculados al sistema.
- Son más limitados en alcance que un registro estadístico, pero dentro de su acotado alcance pueden tener mayor cobertura de unidades y/o variables.
- Contienen una o más variables que no se encuentran en los registros estadísticos del sistema. Tales variables son generalmente usadas con fines de estratificación.
- Son herramientas para incorporar datos administrativos que sólo son relevantes para un subconjunto de unidades en un registro estadístico.
- Pueden contener unidades adicionales, o variables, o ambas cosas.
- Pueden construirse utilizando información de fuentes administrativas, encuestas por muestreo o una combinación de ambas.
- En algunos casos pueden agregar, combinar o transformar variables, aunque en otros casos pueden ser más o menos idénticos a una fuente particular.
- Para garantizar que los registros satélite sean suficientemente coherentes con los registros estadísticos, puede ser útil considerar criterios adicionales, como identificadores comunes de unidades, definiciones y clasificaciones comunes. Cuanto mayor sea la coherencia, más útil será el registro satélite.
- No son registros satélite las bases de datos o archivos de datos en los cuales son normalmente almacenados los resultados de las encuestas por muestreo.

8.3. Transformación e integración de RA en RE

El uso estadístico de datos administrativos requiere de una conversión del registro administrativo a un registro estadístico, como se ha mencionado anteriormente.

El proceso de transformación e integración de registros administrativos en registros estadísticos (creación del registro estadístico) implica una serie de etapas o sub-procesos, como ilustra la figura 18, que son descritos en detalle en los siguientes apartados.

Figura 18. Proceso de transformación de registros administrativos a registros estadísticos.



8.3.1. Controles de consistencia y calidad y depuración de datos

Los registros administrativos no han sido diseñados con fines estadísticos y en muchos casos no aplican las buenas prácticas, recomendaciones o estándares metodológicos requeridos para producir estadísticas.

Los INE no tienen control sobre el proceso de captura de datos de los registros administrativos y es poco frecuente encontrar documentación o los INE tienen poco conocimiento sobre este proceso llevado a cabo por las fuentes administrativas.

Los procesos de ingreso de datos, validación y control de consistencia de la información no siempre aplican los mismos criterios que utilizarían los INE, o las validaciones se aplican con mayor rigurosidad sobre variables administrativas (de interés para la fuente administrativa) que ciertas variables de interés estadístico (utilizadas por los INE) que gozan de menor severidad en sus controles, pues revisten menor importancia para la fuente administrativa.

Además, en cualquier proceso de captación de datos, ya sean registros administrativos, censos o encuestas, se generan errores tanto en los datos que reporta el informante, como quien registra (encuestador o administrativo, si corresponde), en el ingreso de datos (registro en el formulario, digitación, escaneo, etc.), o en la depuración de datos.

Todos estos elementos provocan errores en los datos capturados en los registros administrativos, por lo que es necesario implementar controles de consistencia y calidad para detectarlos y proceder con el proceso de depuración de datos para minimizarlos.

Los errores pueden ocurrir tanto en las variables como en las unidades estadísticas, por lo tanto los controles de calidad y consistencia de los datos deberán aplicarse a ambos niveles.

A partir de las definiciones (metadatos) de las variables estandarizadas (ver siguientes apartados) se deben establecer en un documento los controles de consistencia y calidad (reglas de validación o consistencia) y definir el plan de depuración de los datos. Los resultados de la validación y depuración también deben ser documentados.

8.3.2. Variables del sistema

8.3.2.1. Categorías de variables

Los registros estadísticos están conformados por diferentes categorías de variables. Es importante comprender su utilidad y relevancia dentro del sistema de registros estadísticos.

- Variables **clave o llave de identificación** o variables identificadoras, como su nombre lo indica, se usan para identificar objetos o unidades. Estas variables se usan para hacer la unión entre registros (método determinístico) con el mismo tipo de objetos o unidades. En general se trata de variables numéricas que representan un código de identificación, pero también pueden ser claves alfanuméricas. En el apartado 8.3.3 se presentan diferentes métodos de unión de registros (métodos probabilísticos) cuando no se dispone de una clave identificadora estandarizada o es de mala calidad y se debe recurrir a otras variables como nombres y direcciones.
- Variables de **contacto** (identificadores explícitos) como nombre, dirección, teléfono, correo electrónico se usan cuando el INE necesita contactarse con el objeto o unidad al que corresponde el caso (en general cuando se utilizan cuestionarios). Pero también pueden utilizarse como llaves para la unión de registros por métodos probabilísticos, como se ha mencionado en el párrafo anterior.
- Variables de **ubicación geográfica**, utilizadas para asociar los objetos a ubicaciones físicas en el territorio. Entran en esta categoría las coordenadas geográficas determinadas por GPS y las variables codificadas correspondientes a nomenclátors de divisiones político-administrativas-geográficas del territorio de cada país.
- Variables de **unión o fusión** (claves foráneas), son utilizadas para hacer la unión con otras tablas de una base de datos, describen relaciones entre diferentes tipos de objetos. A diferencia de las claves o llaves de identificación, que tienen una correspondencia uno a uno cuando se hace la unión entre registros (salvo que existan duplicados en alguna de las tablas de los registros), las claves foráneas mantienen una relación uno a muchos. Por ejemplo, supongamos que la tabla correspondiente al registro de población tiene una variable de fusión o clave foránea para

determinar en qué vivienda reside cada persona, esta variable está vinculada con la variable llave de identificación del registro de viviendas. Entonces la tabla del registro de viviendas tiene una relación uno a muchos con la tabla del registro de personas, a través de la variable código de vivienda, pues en una vivienda residen muchas personas¹⁰.

- Variables **estadísticas**, son las variables provenientes del registro administrativo de interés estadístico, que se utilizan para generar estadísticas o hacer análisis estadístico. Se clasifican en variables cuantitativas, continuas (escalares), como por ejemplo edad, ingresos, superficie, ventas, etc.; o variables cualitativas, categóricas (nominales u ordinales), como estado civil, sexo, tipo de vivienda, etc. También entran en esta categoría las variables de estratificación y de ubicación geográfica, que pueden ser a su vez clasificadas como variables estadísticas cuantitativas o cualitativas, como por ejemplo el código compuesto de ubicación geográfica de los inmuebles, código de rama de actividad, ingresos de las personas, volumen de ventas de las empresas o el personal ocupado.
- Variables de **uso administrativo** interno, se utilizan para indicar resultados de procesos internos como por ejemplo códigos de error de las variables estadísticas, valores imputados, resultados de la depuración de variables, código de la fuente administrativa de las variables, código de la fuente o versión del código de clasificación de las variables codificadas (CIUO-08, CIIU-Rev4, etc.), fecha de actualización de las variables, estado de los objetos o unidades (personas: Activo, Inactivo por emigración, Inactivo por muerte, Eliminado por proceso administrativo, Inactivo por duplicado; viviendas: Activa, Inactiva por demolición, Eliminada por proceso administrativo, Inactiva por duplicado), entre otros.
- Variables de **referencia temporal**, indican el momento en que ocurre un evento relativo a los objetos o unidades del registro. Por ejemplo, la fecha de creación de una empresa (inscripción en el registro de la oficina de impuestos, por ejemplo), fecha de construcción/habilitación del inmueble, período al que corresponden los ingresos de una persona (ingresos del último año o del mes anterior), fecha de nacimiento, fecha de cambio de residencia de las personas. Se utilizan para generar diferentes versiones del registro estadístico de acuerdo a los períodos establecidos por esta variable, (por ejemplo, stock de viviendas en determinado período) y para describir la evolución de determinados fenómenos a lo largo del tiempo.
- Variables **derivadas** o **agregadas**, son variables que no están presentes en ningún registro administrativo pero son necesarias para facilitar la

¹⁰ Una misma persona podría ser propietario o inquilino de más de una vivienda, pero para el ejemplo se ha considerado la residencia habitual (en el caso de personas que residen en varias viviendas, se considera una sola de ellas).

producción de estadísticas. El apartado 8.3.2.3 describe el proceso de creación de variables derivadas y las diferentes formas de hacerlo.

8.3.2.2. Estandarización de variables

Las variables que forman parte de los registros administrativos no han sido concebidas con fines estadísticos, por lo tanto en la mayoría de los casos no se ajustan a los conceptos y definiciones necesarios para su uso con fines estadísticos.

Las fuentes administrativas responsables de los registros administrativos se preocupan de que las variables de dichos registros cumplen los propósitos administrativos para los cuales han sido diseñadas, pero en general, los instrumentos de captura de datos utilizados no siguen los estándares, recomendaciones o buenas prácticas de los instrumentos diseñados con fines estadísticos. Es decir, las categorías de respuesta de variables no se ajustan a estas recomendaciones, o los clasificadores/codificadores utilizados para asignar códigos a las variables abiertas no están estandarizados como los habitualmente utilizados por los INE.

Las definiciones de las variables de RA se ajustan a las necesidades administrativas y no siempre se corresponden con las definiciones estadísticas de acuerdo a un uso estadístico específico.

Es así que resulta imprescindible estandarizar las variables del RA para incorporarlas al sistema de registros estadísticos y poder utilizarlas en diferentes proyectos estadísticos.

La estandarización de variables del registro corresponde al INE, quien es responsable de los metadatos de las variables (denominación, definición, categorías, calidad y documentación).

El nombre de la variable debe ser único dentro del sistema de registros estadísticos.

Las variables estandarizadas no deberán ser modificadas por otras áreas del INE, puesto que cualquier cambio en la denominación o definición de las variables estandarizadas podría tener un impacto en otros proyectos estadísticos que estén utilizando dichas variables, provocando además que dejen de ser variables estandarizadas.

El hecho de contar con variables estandarizadas minimiza errores, evita duplicidad de esfuerzos de documentación (quienes utilizan las variables estandarizadas no necesitan volver a documentarlas) y facilita su utilización.

Además, la estandarización de variables a través de sus metadatos, facilita la comparación con variables de nuevos registros administrativos que se deban incorporar al sistema. Mediante esta comparación entre definiciones, categorías y documentación es posible especificar el procedimiento necesario para la

transformación de las nuevas variables en variables estadísticas estandarizadas o la creación de variables agregadas, que también estarán estandarizadas.

Se recomienda utilizar el estándar de metadatos DDI para la documentación de las variables estandarizadas. A continuación se presenta un ejemplo de documentación (ficha de metadatos) de una variable estandarizada en un sistema de registros estadísticos.

FICHA DE METADATOS	
NOMBRE DE LA VARIABLE ESTADO_CIVIL	
DEFINICIÓN Estado civil o conyugal actual de las personas.	
TIPO DE DATOS Variable Numérica – Discreta.	
UNIVERSO DE LA POBLACIÓN El universo está conformado por la población civil no institucional residente en el territorio nacional.	
Categorías	
Valor	Categoría
5	Soltero (a)
2	Casado (a)
1	En unión libre
4	Separado (a) o divorciado (a)
3	Viudo (a)
Preguntas e instrucciones	
PREGUNTA TEXTUAL 6. Actualmente está:	
1 En unión libre 2 Casado (a) 3 Viudo (a) 4 Separado (a) o divorciado (a) 5 Soltero (a)	
INSTRUCCIONES PARA EL INFORMANTE O QUIEN CAPTURA LOS DATOS El objeto de esta pregunta es conocer el estado civil o conyugal actual de las personas.	
Marque X en la alternativa correspondiente al estado conyugal o civil actual de la persona, de acuerdo con las siguientes definiciones, las cuales usted debe explicar a su informante en caso necesario.	
Unión libre. Una persona está en unión libre cuando establece con otra persona del sexo opuesto una relación que implica cooperación económica y cohabitación residencial y sexual, la cual se mantiene única y exclusivamente por voluntad propia sin que estén presentes, vínculos legales.	
Casado(a). Se considera casada aquella persona que ha establecido con otra del sexo opuesto una relación aprobada por la ley, que implica cooperación económica y cohabitación residencial y sexual.	
Viudo(a). Es la persona que habiendo fallecido su cónyuge o compañero (a), no ha vuelto a casarse ni vive en unión libre.	
Separado o divorciado(a). Persona cuya unión (casado o en unión libre) ha sido disuelta por vía legal o de hecho y no se ha vuelto a casar, ni vive en unión libre.	
Soltero(a). Persona que nunca ha contraído matrimonio y que no vive ni ha vivido en unión libre.	

Para diligenciar esta pregunta tenga en cuenta que:

Cuando se trate de madres o padres solteras(os), sondee esta información y trate de establecer si tuvo una unión de carácter estable por un determinado tiempo o no. En caso afirmativo considérela (o) como separada (o), en caso contrario será soltera (o).

No existen criterios de tiempo y espacio de convivencia para definir el estado civil de una persona por lo tanto debe respetarse la información que suministre el encuestado. Por ejemplo:

Pueden encontrarse personas casadas que temporalmente se encuentran separadas por razones económicas, de viaje, enfermedad etc.

Pueden encontrarse personas que a pesar de convivir en una vivienda y pertenecer a un mismo hogar se encuentran separadas.

A los hombres menores de 14 años y mujeres menores de 12 años se les debe marcar la alternativa 5 (soltero) sin necesidad de responder la pregunta.

FUENTE DE INFORMACIÓN

Variable P6 (Estado Civil) del archivo SISBEN1 del registro administrativo SISBEN (Sistema de Selección de Beneficiarios Para Programas Sociales) de la fuente administrativa: DNP – Departamento Nacional de Planeación.

RECODIFICACIÓN Y DERIVACIÓN

Categoría	Código Original	Código Estandarizado
En unión libre	1	5
Casado (a)	2	2
Viudo (a)	3	1
Separado (a) o divorciado (a)	4	4
Soltero (a)	5	3

Es importante mantener copias de los formatos o cuestionarios de captura de datos, manuales y otro tipo de documentos utilizados por la fuente administrativa para capturar la información del registro administrativo. Asimismo, es recomendable transcribir las instrucciones de llenado de los formatos de captura de datos.

La fuente de información también debería registrarse en la ficha de metadatos de las variables estandarizadas, indicando el nombre de la variable o variables administrativas originales utilizadas para generar la variable estandarizada, nombre del registro al cual pertenece(n) y la fuente administrativa.

En el caso que la variable estandarizada requiera de algún proceso de transformación, recodificación o creación (variables agregadas) a partir de una o más variables del registro administrativo, deberá indicarse en el apartado de "Recodificación y derivación".

La estandarización de variables implica además, la transformación de formatos de datos y en algunos casos los datos en sí mismos. Por ejemplo, la variable fecha de nacimiento de un registro administrativo almacena los datos en el formato dd/mm/aaaa, pero la variable estandarizada fecha de nacimiento del registro estadístico de población se ha definido con el formato aaaa-mm-dd, lo cual implica que los formatos de datos originales de la variable del registro administrativo deben transformarse (siguiendo un algoritmo) para convertirlos al formato estandarizado de la variable del registro estadístico.

Otro ejemplo se puede ver en las variables categóricas donde los códigos de categoría y sus descripciones podrían cambiar al transformarlos a las correspondientes variables estandarizadas del registro estadístico.

Supongamos que una variable categórica del registro administrativo como tipo de material de los pisos de la vivienda tiene las siguientes categorías de respuesta: 1-Tierra; 2-Arena; 3-Cemento; 4-Madera o tabla; 5-Otro material. Pero su correspondiente variable estandarizada en el registro estadístico tiene las siguientes categorías estandarizadas: 1-Tierra o arena; 2-Cemento; 3-Madera o tabla; 4-Otro material. El proceso de transformación o estandarización implica que los códigos de la variable sean convertidos de la siguiente manera: 1 y 2 ->1; 3->2; 4->3; 5->4. Los códigos originales de la variable administrativa son 1 y 2 que se han unido en un solo código (1) en la variable estandarizada y eso provoca que también el resto de los códigos deban ser transformados.

El INEC de Ecuador ha desarrollado una herramienta de gestión de metadatos denominada Metadec que permite documentar esta asociación entre los catálogos o categorías de las variables. Podría ser un buen punto de partida para desarrollar un sistema completo de documentación de metadatos del sistema de registros estadísticos.

Todos estos procesos de transformación o estandarización deben ser documentados en la ficha de metadatos de las variables estandarizadas.

8.3.2.3. Variables derivadas o agregadas

En la generación de estadísticas basadas en registros administrativos no se tiene la posibilidad de diseñar los instrumentos de captura de datos, no es posible definir las preguntas del cuestionario. En estos casos se debe apelar a la creación de variables agregadas utilizando las variables disponibles de los registros administrativos.

Las variables derivadas o agregadas se crean de seis modos diferentes:

1) Por medio de **cálculos aritméticos** o **procedimientos lógicos**.

En el caso de las variables cuantitativas se pueden hacer operaciones aritméticas o lógicas para generar variables derivadas. Las variables cualitativas sólo permiten hacer operaciones lógicas. Por ejemplo, el área total construida de un inmueble se calcula como la suma de la superficie de cada una de las plantas.

2) Mediante **agrupamiento** de valores.

Se crean agrupando valores en determinados intervalos. Por ejemplo, la ocupación de una persona se podría agrupar en una nueva variable por clases de ocupación.

3) Mediante **codificación**.

En el caso de variables cuyos datos están almacenados en forma de texto no estructurado pues corresponden a preguntas abiertas, como por ejemplo ocupación, rama de actividad, estudios cursados, etc., es necesario convertirlas a datos estructurados mediante un procedimiento de codificación.

4) Por **asociación** de otras variables.

Se calcula la variable derivada usando variables de otro registro. Por ejemplo, se pueden vincular o asociar variables de la vivienda a las personas como ser el tipo de vivienda donde reside cada persona, uniendo ambos registros (personas y viviendas) a través de variables (clave) comunes.

5) Mediante **agregación** de otras variables.

La variable derivada se genera en un registro usando variables de otro(s) registro(s). Se utilizan operaciones aritméticas (suma, promedio, etc.) para agregar los valores de la variable del segundo registro (vinculado) y el resultado se almacena en la variable derivada para cada objeto vinculado. Por ejemplo, una variable derivada que tenga el promedio de horas trabajadas por las personas en cada empresa; o el ingreso total de los hogares o viviendas a partir de los ingresos de las personas que las integran.

6) Mediante **modelos estadísticos**.

Se diseña un modelo estadístico mediante el cual se analiza la relación entre la variable derivada que se desea crear y las variables administrativas, y permite crear la variable en cuestión.

Las variables derivadas se pueden generar combinando estos métodos, incluso aplicando cálculos más complejos con procedimientos basados en ciertas reglas.

Las variables derivadas o agregadas también deben ser documentadas de acuerdo al estándar de documentación definido para tales fines. Se debe hacer especial hincapié en describir detalladamente la fórmula o regla de cálculo utilizada para crear la variable.

8.3.2.4. Trazabilidad de variables

A efectos de disponer de una adecuada documentación de metadatos es fundamental mantener la trazabilidad de cada variable que conforma el registro estadístico, desde su origen en el registro administrativo, donde fue creada por la respectiva fuente administrativa, pasando por el proceso de validación, depuración y estandarización y justificación de su selección e inclusión en el registro estadístico.

El sistema de documentación de metadatos debe garantizar la trazabilidad de todas las variables del sistema de registros estadísticos con sus respectivas

fuentes administrativas (variables administrativas utilizadas para generar las variables estadísticas).

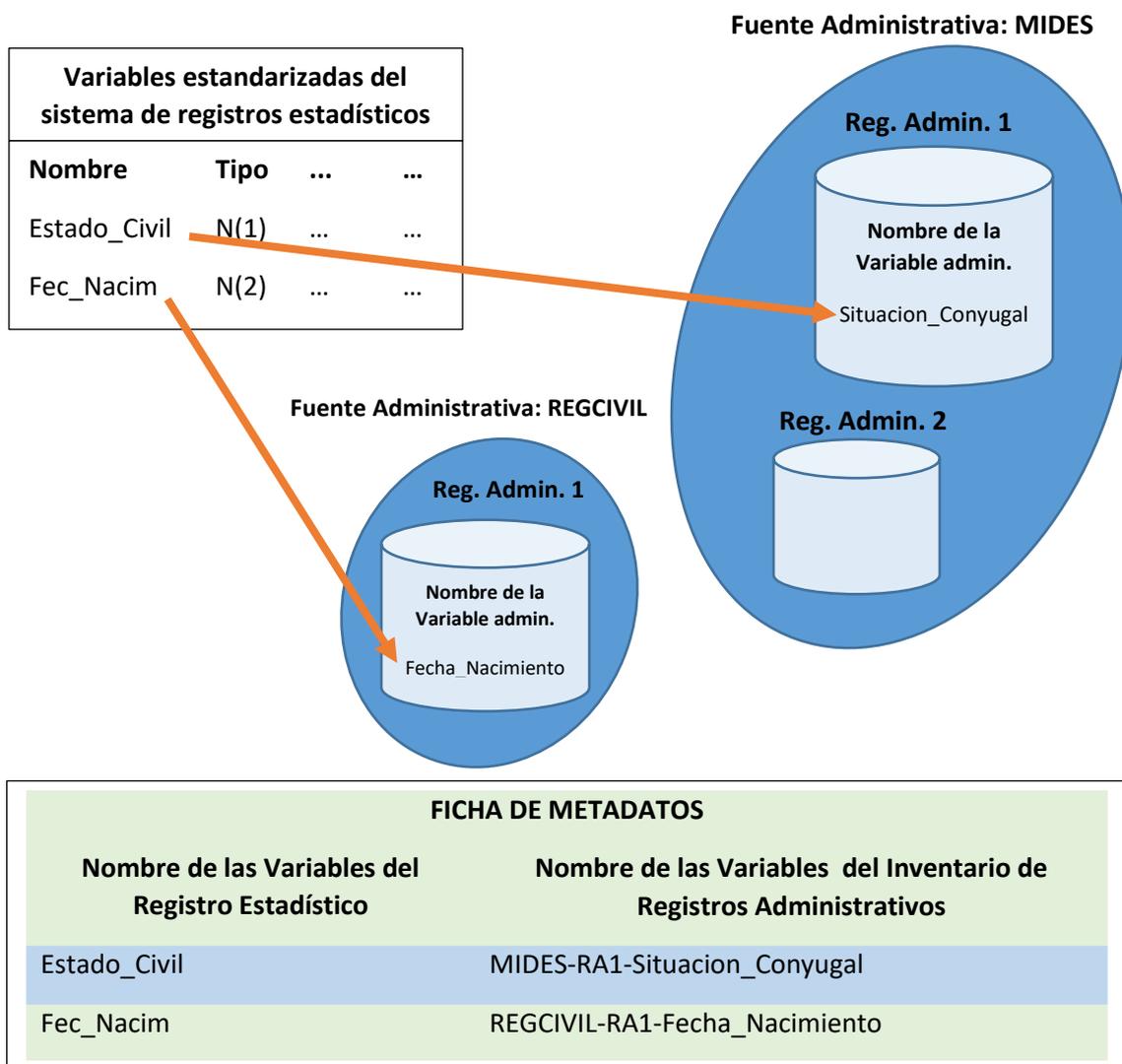
Debe existir una vinculación entre el inventario de registros administrativos y los metadatos del sistema integrado de registros estadísticos.

El nombre de la variable (único dentro del sistema de registros estadísticos, como se ha mencionado anteriormente) es la clave de vinculación con las variables administrativas. Estas variables administrativas deben figurar en el inventario de registros administrativos y deben estar documentadas de acuerdo al estándar de metadatos DDI.

La clave de vinculación de las variables administrativas del inventario de registros administrativos está compuesta por el código de identificación de la fuente administrativa, seguido del código de registro administrativo más el nombre de la variable administrativa (la clave es un código compuesto ya que el nombre de la variable podría repetirse en diferentes registros administrativos).

La siguiente figura ilustra la forma de documentar la trazabilidad de las variables del sistema de registros estadísticos hasta llegar a las variables administrativas que las originan.

Figura 19. Trazabilidad de variables del sistema integrado de registros estadísticos.



Fuente: elaboración propia del consultor.

Las variables del registro estadístico también pueden conformarse combinando varias variables, por medio de fórmulas (que son especificadas en la ficha de metadatos en el apartado de “Recodificación y Derivación”), cuyas fuentes pueden ser otras variables administrativas o estadísticas, o la combinación de ambas.

A continuación se presentan algunos ejemplos:

FICHA DE METADATOS		
Nuevas Variables del Registro Estadístico	Origen de los datos	
	Variables del Inventario de Registros Administrativos	Variables del Registro Estadístico
Estado_Civil	MIDES-RA1-Situacion_Conyugal	
Fec_Nacim	REGCIVIL-RA1-Fecha_Nacimiento	
Ingresos_Persona	SegSocial-APORTES-Salario; SegSocial-APORTES-HrsExtra; SegSocial-PENSIONES-Pension	
Egresos_Persona	DGI-DeclImpuestos-Cuota_Coche	Engresos_Persona

En la ficha de metadatos debe justificarse por qué se eligió determinada(s) fuente(s) para crear cada variable del registro estadístico. Las justificaciones estarán basadas en los reportes de la calidad de los registros administrativos y en la evaluación de la calidad del registro estadístico.

8.3.3. Unión de registros

Según Fellegi y Sunter *record linkage* o unión de registros "es una solución al problema de identificar aquellos registros (filas o casos) en dos archivos que representan personas, objetos o eventos idénticos" ¹¹.

La unión de registros es la tarea de identificar de forma rápida y precisa los registros correspondientes a la misma entidad/objeto/individuo de una o más fuentes (archivos) de datos ¹².

A lo largo de este documento se ha mencionado la necesidad de unir registros de diferentes fuentes administrativas para integrarlos y transformarlos en registros estadísticos (y entre registros administrativos y registros estadísticos), para lo cual es necesario contar con claves únicas de identificación y en algunos casos claves foráneas también.

En un mundo ideal, o más parecido a la realidad de los países nórdicos, todos los registros administrativos deberían contar con una variable clave de identificación de los casos (en general es una única variable, pero hay casos donde se utiliza más de una variable para crear la clave de identificación). Además, esta clave debería estar estandarizada y utilizada por todos los

¹¹ Fellegi I.P., Sunter A.B. (1969) *A theory for record linkage*. Journal of the American Statistical Association 64, 1183-1210. USA.

¹² Baxter, R. Gu, L. Vickers, D. Rainsford, C. *Record Linkage: Current Practice and Future Directions*. CSIRO Mathematical and Information Sciences. CMIS Technical Report No. 03/83. Canberra, Australia.

registros administrativos referidos al mismo tipo de objetos o elementos. De esta forma se podrían unir filas de diferentes registros que refieren al mismo objeto o elemento, en los casos que coincida exactamente la clave de identificación de ambos registros (método determinístico).

Por ejemplo, en el caso del registro civil las personas tienen como clave la cédula de identidad o DNI y esta variable clave es utilizada en todos los otros registros de población como clave de identificación de cada persona. Se busca lograr una coincidencia exacta carácter a carácter de la variable clave, entre ambos registros.

La realidad en la subregión dista bastante de este escenario ideal. En todos los países existen varios casos de registros administrativos que no utilizan una clave única estandarizada común para identificar los casos, sobre todo en los registros de inmuebles.

En las situaciones donde sí se utiliza una clave de identificación común, ésta presenta ciertos problemas, como por ejemplo: contiene duplicados, o la estandarización no es de alcance nacional entonces cada municipio tiene su propio código y procedimiento para crearlo.

Si se aplica el método determinístico de unión de registros sin evaluar la calidad de la variable clave utilizada, se corre el riesgo de unir filas o casos de ambos registros que no se tiene la certeza que correspondan realmente al mismo objeto o elemento. Asimismo, si se tienen duplicados no se sabrá cuál de las filas es la que corresponde efectivamente al elemento u objeto en cuestión. Además, están los casos cuya clave de identificación está en blanco o contienen datos inválidos y no será posible unirlos con otros registros.

Por estas razones se deben plantear métodos alternativos para la unión de registros. Los métodos probabilísticos de unión de registros utilizan algoritmos específicos para determinar con cierta certeza (probabilidad) que dos filas o casos de diferentes registros corresponden al mismo elemento u objeto. Utilizan otras variables del registro (aparte de las variables clave de identificación), combinándolas para lograr una pseudo-clave. Las variables de población más comúnmente utilizadas para estos propósitos son nombre, apellido, fecha de nacimiento o edad y otras dependiendo de la disponibilidad en el registro. En el caso de inmuebles se utiliza la variable de dirección del inmueble.

Las variables alfanuméricas que podrían ser utilizadas para hacer la unión de registros presentan una serie de problemas:

- Errores, variaciones y datos en blanco.
- Diferencias en definiciones, períodos, formatos de los datos capturados por los diferentes registros administrativos.
- Cambios en los datos a lo largo del tiempo, por ejemplo cambio de domicilio de las personas.
- Errores de digitación, letras o palabras ingresadas en diferente orden.

- Palabras fusionadas o divididas, palabras incompletas, letras faltantes o excedentes.
- Puntuación o acentuación incorrecta.
- Abreviaciones.

Por lo tanto es imprescindible realizar la depuración de datos y estandarización de variables antes de iniciar el proceso de unión de registros.

Los métodos de unión probabilística de registros implican el cálculo de pesos de unión estimados con base en todos los casos coincidentes y no coincidentes observados de los valores de la(s) variable(s) utilizada(s) para la unión.

Los métodos probabilísticos permiten obtener una mejor unión de registros que un simple método de unión determinístico. Además, pueden ser utilizados para detectar casos/filas duplicados en un archivo de datos del registro administrativo (cuando dos o más casos/filas tienen diferentes valores en la clave de identificación, pero en realidad corresponden a la mismo objeto/elemento/individuo).

El proceso de unión probabilística de registros consta de tres fases:

- 1) **Pre-unión.** Esta etapa implica la depuración de datos y estandarización de variables, como se ha explicado en apartados anteriores.

Además, en el caso que se utilicen variables alfanuméricas para realizar la unión de registros, éstas presentan una gran cantidad de errores tipográficos o problemas de conversión de caracteres especiales, por lo cual deben normalizarse, es decir, los textos deben ser procesados para homogeneizarlos de forma tal que los métodos de unión de registros utilizados logren una mayor tasa de coincidencias.

Para esto, se deberían aplicar los siguientes criterios de normalización de textos:

- Reemplazar las ocurrencias del símbolo ¥ (o el símbolo correspondiente al sistema de codificación utilizado en las bases de datos o archivos) por el carácter Ñ, y el símbolo ◆ por la ñ ¹³.
- Sustituir las vocales con acento o diéresis por las respectivas vocales.
- Se convierten las minúsculas a mayúsculas.
- Eliminar partículas. (D, DE, DEL, DA, DI, DO, L, LA, LAS, EL, LOS, Y)
- Reemplazar los ceros por la letra O. Excepto en la variable de direcciones.
- Eliminar los números. (1, 2, 3, 4, 5, 6, 7, 8, 9). Excepto en la variable de direcciones.

¹³ Los símbolos ¥ y ◆ podrán variar dependiendo del sistema de codificación o set de caracteres utilizado (UTF-8, ISO-8859, etc.).

- Se eliminan los caracteres diferentes a las 27 letras mayúsculas y minúsculas del alfabeto. Se eliminan los guiones, puntos, comas y otros caracteres.

Las variables de domicilios son un caso particular de variables alfanuméricas y se deben normalizar siguiendo los criterios de estandarización de cada país. En el caso que los domicilios se almacenen en más de una variable (calle, número, complemento, piso, apartamento, etc.) éstas se deberían concatenar en una sola variable siguiendo los mismos criterios de estandarización.

2) **Unión.** Se aplica la unión de registros para decidir cuando dos casos o filas de diferentes registros coinciden (match), o sea pertenecen al mismo objeto o elemento, o no coinciden (no-match) es decir, se trata de diferentes objetos o elementos. También puede utilizarse para encontrar duplicados en el mismo archivo del registro administrativo. Se utilizan herramientas informáticas que proveen varios algoritmos para realizar la unión probabilística. Estas aplicaciones primero calculan los pesos para cada uno de los posibles casos de unión y luego se determinan los umbrales de los casos unidos y los no unidos. A modo de resumen, uno de los métodos más conocidos y comúnmente aplicado es el método Fellegi-Sunter que sigue estos pasos:

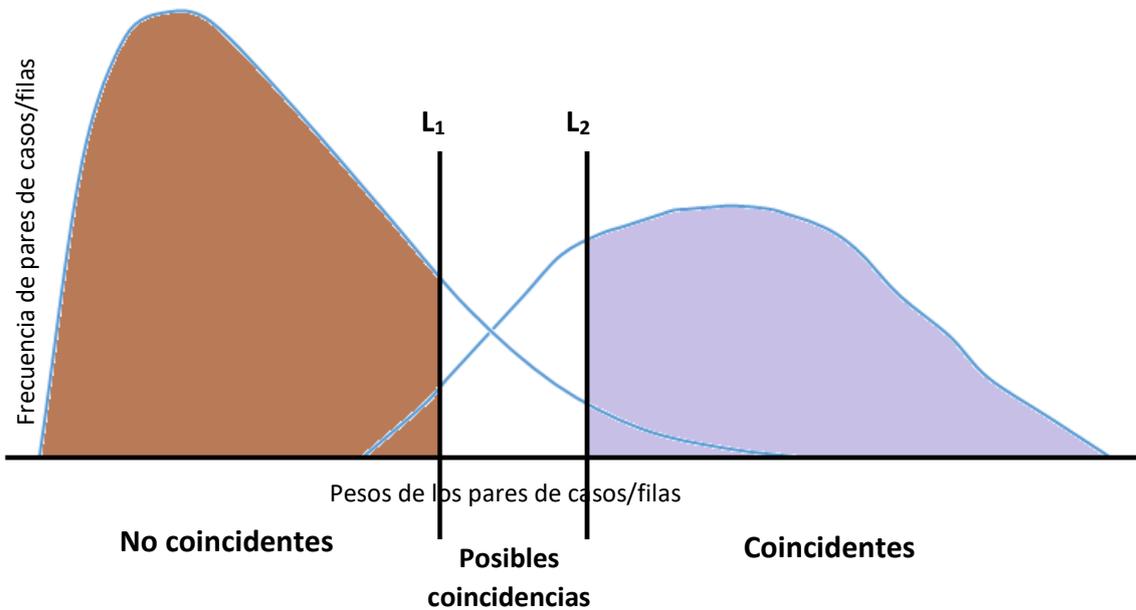
- Cada variable utilizada en el proceso de unión es comparada y se le asigna un puntaje (peso) basado en qué tan bien coincide.
- Se calcula un puntaje para cada campo (variable) que indica, para cada par de casos/filas, cuan probable es que ambos correspondan al mismo objeto o elemento. Este puntaje es basado en la probabilidad de que la coincidencia entre los campos corresponda a una verdadera coincidencia de los objetos o elementos.

El método

- Se suman los puntajes de todos los campos. El puntaje total de la unión entre dos casos/filas de dos registros es la suma de los puntajes generados de la unión de los campos individuales.
- Se ordenan los pares de casos/filas de acuerdo a sus puntajes (pesos).
- Se establecen ciertos valores de corte de los puntajes para distinguir entre coincidencias (matches) y no coincidencias (no-matches).
- Por encima de cierto umbral de puntaje (peso) se considera como coincidencia (match).
- Por debajo de cierto umbral se considera como no coincidencia (no-match).
- Todo lo que se encuentre entre ambos umbrales se considerará como “posible coincidencia” que necesita ser revisada manualmente (si es que se elige esta alternativa). La siguiente figura ilustra cómo los dos límites o umbrales dividen los pares de casos/filas en tres zonas. La zona de la izquierda del límite L_1 corresponde a los pares de casos/filas clasificados como no coincidentes, en la zona derecha del límite L_2 se encuentran los pares de casos/filas catalogados como coincidentes y en la zona del

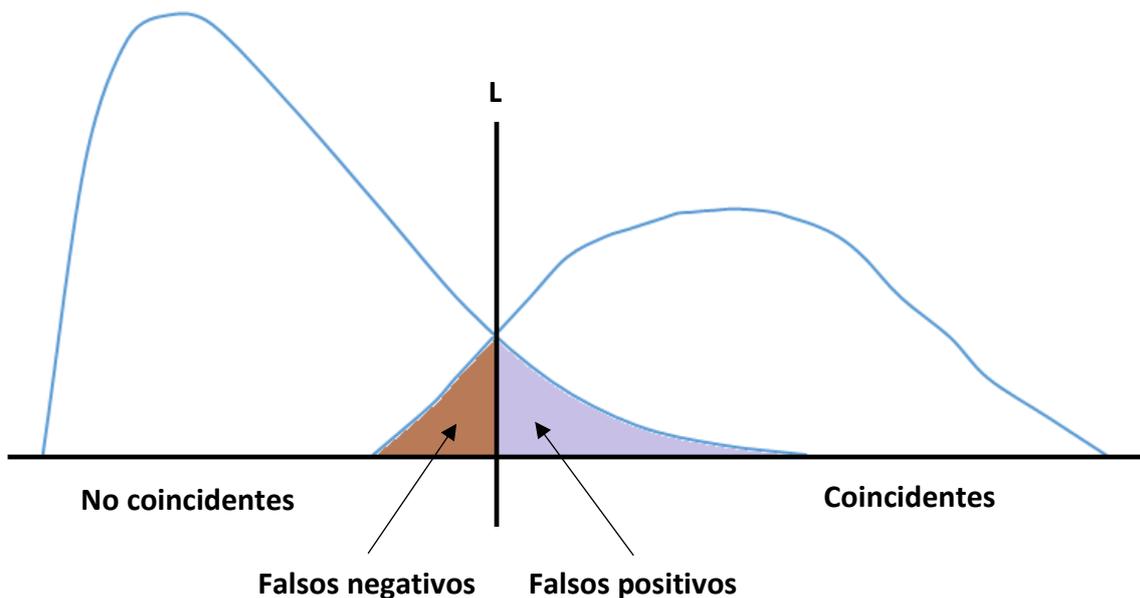
medio entre los límites L_1 y L_2 se hallan los pares de casos/filas denominados “posibles coincidencias” y los cuales deberán pasar por un proceso de revisión manual para decidir en qué grupo clasificarlos.

Figura 20. Umbrales que delimitan las zonas correspondientes a los tres grupos de pares de casos/filas: no coincidentes, posibles coincidencias y coincidentes.



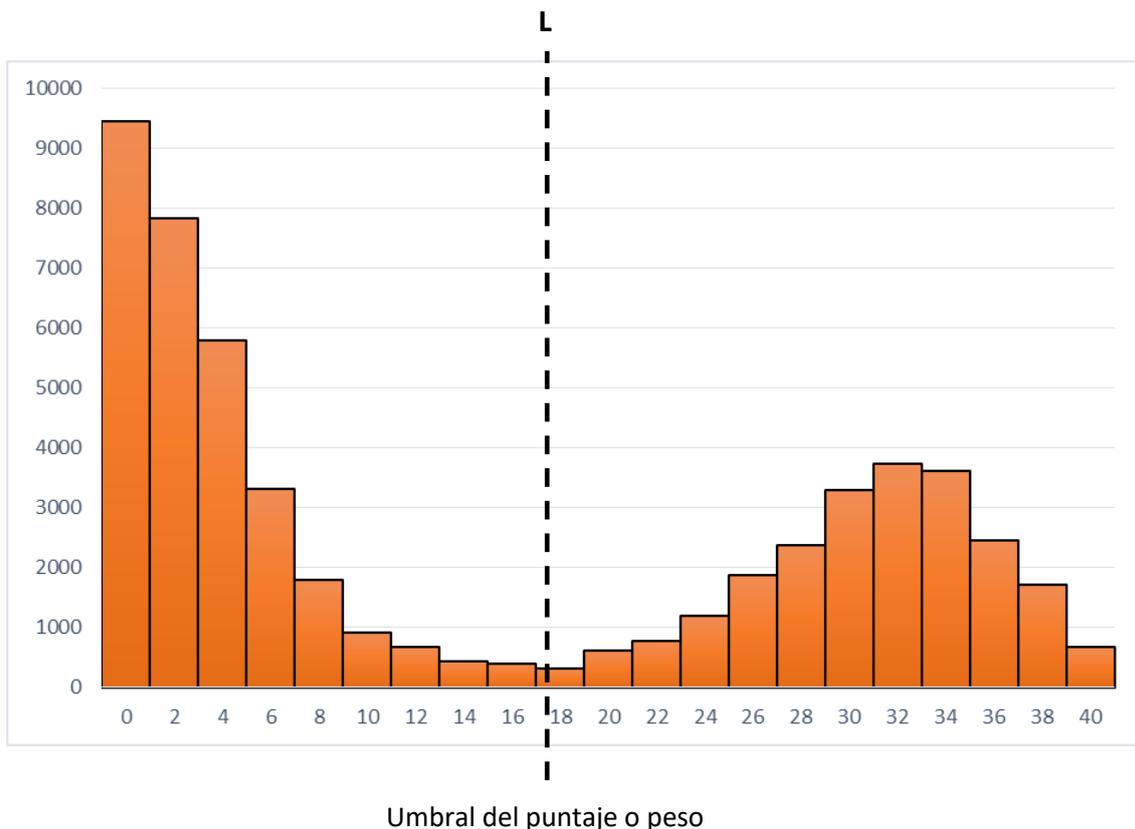
- Otra opción es seleccionar un único umbral de corte (L) y todos los puntajes que estén por encima serán considerados como coincidencias y los que estén por debajo de ese mismo umbral serán calificados como no coincidentes, sin dejar lugar a “posibles coincidencias” (ver siguiente figura), es decir, todo el proceso será automático.

Figura 21. Fijación de un único umbral de corte para dividir en dos grupos de pares de casos/filas: no coincidentes y coincidentes.



- La forma más simple de determinar adecuadamente este límite (L) es mediante la observación del histograma de frecuencias de todos los pesos totales calculados para todos los pares de casos/filas. La forma que presenta el histograma ayudará a tomar la decisión acerca de dónde ubicar el umbral (L). El siguiente gráfico representa un histograma de frecuencias de todos los pesos totales asignados a los casos/filas.

Gráfico 1. Histograma de frecuencias de los pesos totales de los pares de casos/filas, utilizado para la fijación de un único umbral de corte para dividir los pares en dos grupos: no coincidentes y coincidentes.



Los pares de casos/filas correspondientes al mismo objeto o elemento tienen, en general, un mayor peso, mientras que los pares que no corresponden al mismo objeto, elemento o individuo deberían tener un menor peso.

- Es así que mediante la observación del histograma de frecuencias se puede establecer el valor del umbral (L) y todos aquellos pares de casos/filas que tienen un peso superior a este límite se catalogan como coincidentes y aquellos con pesos por debajo del límite serán considerados como no coincidentes.

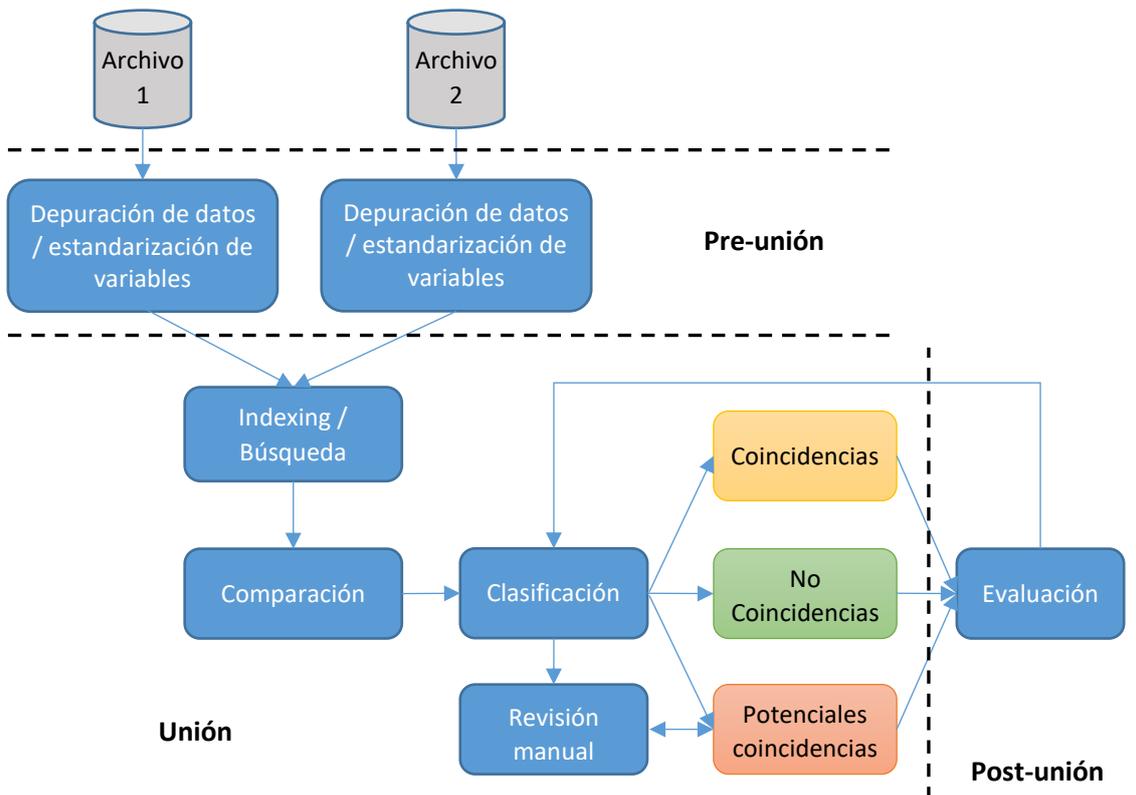
Otros métodos probabilísticos de unión de registros son Damerau-Levenshtein, Needleman Wunsch, Jaro-Winkler, Pair letters Similarity, Metaphone, SoundEx.

Además, el DANE de Colombia ha desarrollado dos algoritmos con buenos resultados: comparación de letras (se concatenan los nombres y apellidos de una persona, se cuenta el número de veces que aparece cada letra y se almacena una cadena de 56 dígitos; se determina si hay coincidencia si hay al menos 10 dígitos iguales e igual fecha de nacimiento; se logró un 95% de concordancia) y método fonético (a cada nombre se le asigna el equivalente fonético y se realiza el emparejamiento por medio de estos campos fonéticos junto con la fecha de nacimiento).

3) **Post-uniión.** Revisión manual de los casos/filas no vinculados (si se opta por esta estrategia). Evaluación de los resultados.

El diagrama de procesos de la siguiente figura representa esquemáticamente el proceso de unión de registros.

Figura 22. Diagrama resumen del proceso de unión de registros.



La unión probabilística de registros presenta los siguientes desafíos:

- No hay identificadores únicos.
- Los datos de los registros administrativos contienen errores.
- Archivos de datos demasiado grandes.
- No hay datos de entrenamiento (o testing para hacer ajustes) en muchas aplicaciones de unión de registros.

- Privacidad y confidencialidad.

Para solucionar el problema de manejar archivos de datos demasiado grandes se pueden utilizar técnicas de *blocking*, *indexing* o *filtering*, para trabajar con bloques del archivo de menor tamaño.

La técnica de *indexing* permite fácilmente remover pares de casos/filas que obviamente no coinciden.

El *blocking* solo compara pares de casos/filas que tienen el mismo valor en la variable elegida como variable de blocking. Se recomienda utilizar más de una variable de blocking para minimizar los errores debidos a datos erróneos en dichas variables.

Recientemente se han estado aplicando técnicas de *aprendizaje automático* o *machine learning* para resolver los problemas que presenta la unión de registros.

Los métodos de aprendizaje automático como arboles de decisión, redes neuronales, aprendizaje basado en ejemplos o *instance-based learning*, agrupamiento o *clustering*, entre otros, se utilizan ampliamente para la clasificación de patrones. Un algoritmo de aprendizaje automático construye un modelo a partir de información suministrada en forma de ejemplos, para generalizar comportamientos y reconocer patrones.

Los métodos de aprendizaje automático se clasifican en dos grupos: aprendizaje supervisado (se cuenta con información que especifica qué conjuntos de datos son satisfactorios para el objetivo del aprendizaje) y aprendizaje sin supervisión (encontrar patrones que permitan separar y clasificar los datos en diferentes grupos, en función de sus atributos).

Esta técnica de aprendizaje automático se podría implementar a través de herramientas informáticas específicas para incorporarla en el futuro SIREPI. Por tanto, sería deseable analizar esta opción en el diseño del SIREPI.

Caso particular: el problema de la falta de identificador común entre personas y viviendas, e incluso entre los registros de inmuebles tampoco se cuenta encuentra una clave de identificación estandarizada. En su lugar se deben utilizar las variables de domicilios/direcciones.

En este caso se deben normalizar las variables de domicilio siguiendo los criterios de estandarización de cada país. En el caso que los domicilios se almacenen en más de una variable (calle, número, complemento, piso, apartamento, etc.) éstas se deberían concatenar en una sola variable siguiendo los mismos criterios de estandarización.

Tal como se ha mencionado en el informe de la síntesis del diagnóstico subregional (Anexo I), el registro administrativo de inmuebles que tiene la mayor cobertura y calidad en las variables de domicilios es el registro de servicios

públicos, en especial el de energía eléctrica que incluso cuenta con las coordenadas geográficas de los medidores de consumo. Se debería integrar este registro con otros registros de inmuebles (catastro) para mejorar la cobertura del registro y la calidad de la variable de domicilios.

Luego de aplicar las técnicas de unión probabilística de registros se deberían implementar otros métodos complementarios para lograr una mayor tasa de coincidencias. Se podría contar con el apoyo de herramientas de SIG o GIS y trabajar con capas para determinar la ubicación aproximada de la vivienda a nivel de manzana al menos, utilizando los datos de las coordenadas geográficas de los medidores de energía eléctrica.

8.4. Resumen de pasos para la implementación de la metodología

- 1) Especificar necesidades, establecer objetivos, definir alcance.
- 2) Elaborar plan de acción por etapas (implementación gradual). Establecer objetivos y estadísticas a generar en cada etapa, de modo de priorizar el acceso a los registros administrativos.
- 3) Crear/actualizar inventario de registros administrativos asociados a las necesidades de información (demanda) y objetivos de cada etapa.
- 4) Contactar fuentes administrativas y recibir datos administrativos de los proveedores.
- 5) Definir población, objetos o unidades estadísticas y variables del registro estadístico.
- 6) Documentar los metadatos de los registros administrativos (diccionario de variables) de acuerdo al estándar DDI.
- 7) Evaluar la calidad de los registros administrativos. Aplicar la herramienta HECRA¹⁴ en sus tres primeras dimensiones (fuente administrativa, metadatos y datos).
- 8) Verificar y depurar los datos administrativos entregados por los proveedores de datos. Detectar duplicados. Registrar resultados de la verificación de acuerdo a los instrumentos que forman parte del presente documento.
- 9) Transformar el registro administrativo en registro estadístico. Estandarizar variables, homologar categorías, etc. Documentar el proceso de transformación por medio de los instrumentos que son parte de la presente metodología.

¹⁴ Segui Stagno, Federico (2012). *Guía de la herramienta para la evaluación de la calidad de Registros Administrativos (HECRA) a ser usados con fines estadísticos*. Banco Mundial. México, 2012.

- 10) Integrar diversas fuentes administrativas en el registro estadístico. Aplicar métodos determinísticos y probabilísticos de emparejamiento de registros. Documentar el proceso de integración haciendo uso de los instrumentos aquí desarrollados.
- 11) Aplicar la protección de datos. Desidentificación del registro estadístico. Sustituir las variables clave de identificación por claves aleatorias y eliminar las variables de contacto o identificadores explícitos (nombres, teléfono, dirección, email, coordenadas geográficas).

ATENCIÓN: este proceso de protección de datos simplemente hace la desidentificación de registros. No se trata de un proceso de anonimización completo, el cual debería implementarse antes de liberar una base de microdatos a usuarios externos.
- 12) Crear/actualizar registro base, registro estadístico primario o registro estadístico integrado, según corresponda.
- 13) Crear/actualizar metadatos del registro estadístico (actividad realizada en paralelo a todas las etapas).
- 14) Asegurar la calidad del proceso (actividad realizada en paralelo a todas las etapas). Aplicar herramienta de evaluación de la calidad del sistema integrado de registros estadísticos.
- 15) Retroalimentación a las fuentes administrativas para mejorar la calidad de sus registros administrativos.
- 16) Comunicar las actualizaciones de los registros estadísticos del sistema.

8.5. Adaptaciones a cada país

La metodología propuesta ha sido elaborada de acuerdo al contexto subregional con un enfoque armonizado, sin embargo es necesario realizar adaptaciones a las particularidades de cada país y cada INE, sobre todo en lo que respecta a su implementación.

Estas adaptaciones serán realizadas durante el taller regional de febrero, evento en el cual participarán representantes de cada país y así se podrán efectuar estos ajustes.

8.6. Plan de implementación estándar

En este apartado se presentará un plan de implementación estándar adaptado a cada país. A realizarse durante el taller regional de febrero.

9. Metadatos, herramientas e instrumentos para la documentación

En este capítulo se exponen los modelos y sistemas de documentación de registros estadísticos, herramientas para la documentación sobre la estandarización de variables de los registros estadísticos a partir de registros administrativos, mapeo o trazabilidad de variables y fuentes administrativas, formatos de registro de los resultados de los controles de calidad y consistencia de los datos, estandarización de variables e integración de registros estadísticos.

A continuación se presentan los modelos de metadatos y estándares de documentación más apropiados para cada fase definida por el modelo GSRBPM.

1. Especificar necesidades

- Documento para especificar los requerimientos de los usuarios.
- Ficha del inventario de registros administrativos.
- Convenio estándar de entrega de información de la fuente administrativa.
- Ficha de metadatos del sistema de registros estadísticos. Metadatos del proceso (paradata).

2. Diseño

- Ficha de metadatos en formato DDI para definir población, objetos y variables.
- Documento de diseño de los instrumentos de recolecta de datos, procesamiento, integración, actualización de RE y sistemas de producción y flujos de trabajo.
- Cronograma de trabajo, plan de implementación.
- Ficha de metadatos del sistema de registros estadísticos. Metadatos del proceso (paradata).

3. Desarrollar

- Cronograma de trabajo actualizado.
- Reporte de resultados de las pruebas.
- Ficha de metadatos del sistema de registros estadísticos. Metadatos del proceso (paradata).

4. Recolectar datos

- Cronograma de trabajo actualizado.
- Reporte de evaluación de la calidad de los RA.
- Ficha de metadatos del sistema de registros estadísticos. Metadatos del proceso (paradata).

5. Transformar

- Cronograma de trabajo actualizado.
- Reporte de resultados del control de inconsistencias y depuración de datos.
- Ficha de metadatos del sistema de registros estadísticos. Estandarización de variables. Metadatos del proceso (paradata).

6. Integrar

- Reporte de resultados de la integración.
- Ficha de metadatos del sistema de registros estadísticos. Derivación de variables. Metadatos del proceso (paradata).

7. Comunicar

- Comunicados internos.
- Ficha de metadatos del sistema de registros estadísticos. Metadatos del proceso (paradata).

8. Evaluar

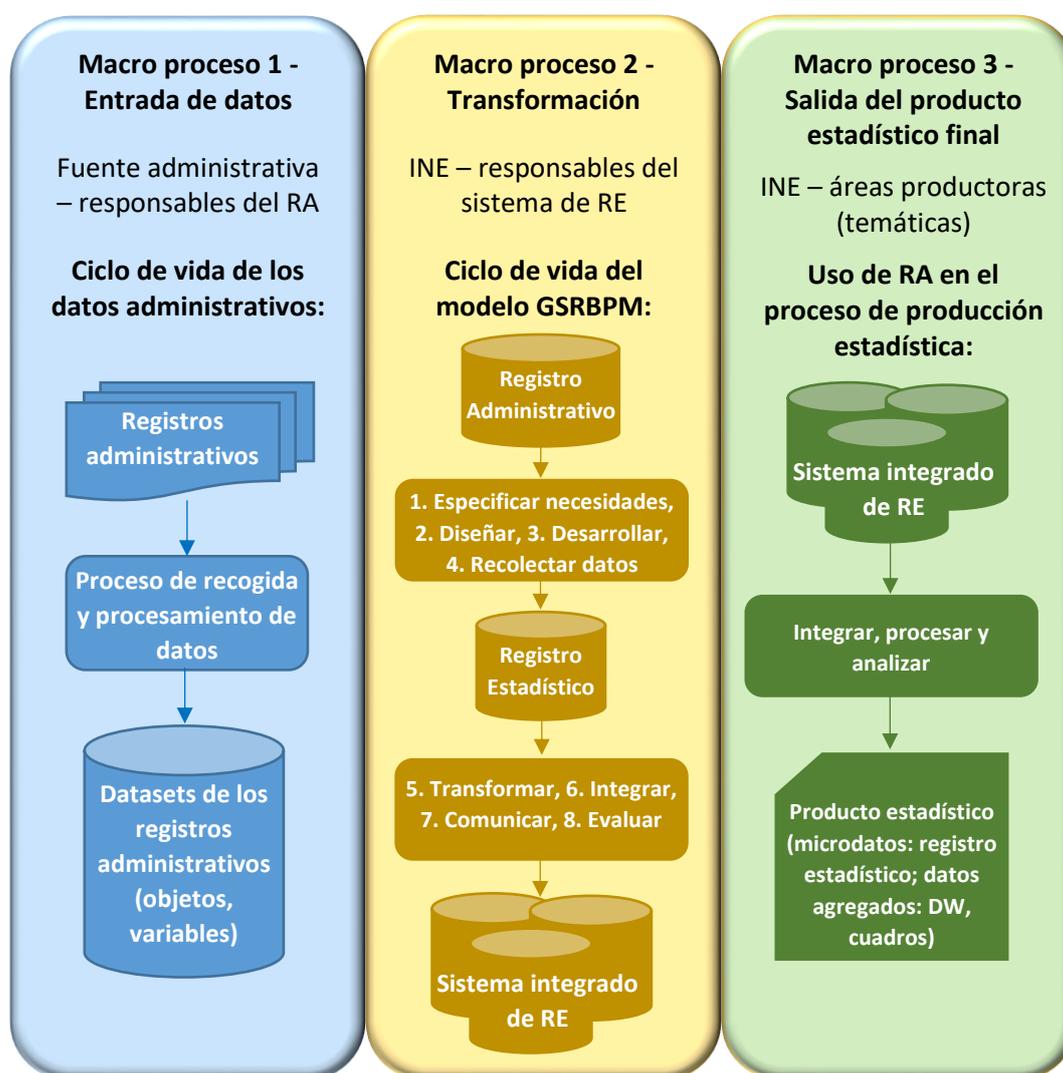
- Reportes de resultados de actividades, reportes de la calidad, evaluaciones de resultados de integración de registros.
- Reporte de la evaluación.
- Plan de mejora acordado.
- Ficha de metadatos del sistema de registros estadísticos. Metadatos del proceso (paradata).

10. Modelo de evaluación de la calidad del sistema integrado de registros estadísticos

El modelo de evaluación de la calidad del sistema integrado de registros estadísticos está basado en la herramienta HECRA del Banco Mundial y en modelos de evaluación de Suecia, Holanda, Italia, Estados Unidos, entre otros.

El modelo propuesto identifica tres macro procesos relativos a la producción de estadísticas basadas en registros administrativos, representados por la siguiente figura. Estos macro procesos son: entrada de datos, proceso de transformación (procesos del modelo **GSRBPM**) y salida del producto estadístico final.

Figura 23. Macro procesos de producción estadística basada en registros administrativos.



Fuente: adaptado de Brancato, G. y otros (2016). *Guidelines for the quality of statistical processes that use administrative data*. Istat. Italia.

Está orientado a evaluar la calidad de entrada (datos administrativos), calidad del proceso de transformación y calidad de salida (producto estadístico)¹⁵. Los elementos de la calidad de entrada a ser evaluados son los atributos propuestos por la herramienta HECRA para la dimensión de *Fuente administrativa de datos*. A los que se le agregarán los atributos de la calidad asociados al proceso de transformación e integración de registros administrativos en estadísticos (dimensiones de la calidad de *Metadatos* y *Datos* de la HECRA) y los elementos de calidad correspondientes al producto estadístico final (dimensión de la calidad “producto estadístico final” de la HECRA).

¹⁵ Adaptado de los modelos propuestos por Zhang L.C. (2012). *Topics of statistical theory for register-based statistics and data integration*; Istat de Italia: Brancato, G. y otros (2016). *Guidelines for the quality of statistical processes that use administrative data*; y Laitila, T., Wallgren, A. Wallgren B. (2011). *Quality Assessment of Administrative Data*.

Un registro administrativo puede ser utilizado de tres (o cuatro) modos diferentes en un sistema de registros estadísticos, Laitila, T., Wallgren, A. Wallgren B. (2011):

- 1) Parea generar productos estadísticos tal cual está en la fuente administrativa.

Si la calidad del registro es adecuada, se puede usar como está para elaborar un producto estadístico. La evaluación de la calidad para determinar si es adecuada se realiza de acuerdo con los indicadores de la calidad del producto estadístico.

- 2) Para generar productos estadísticos luego de ciertos procesamientos y transformaciones.

Se subdivide en dos casos: i) existen algunos problemas de calidad con el RA, pero luego de ciertos procesamientos utilizando otras fuentes, es posible usar el RA para generar el producto estadístico; y ii) el RA ya se utiliza para producir cierto producto estadístico, pero después de un procesamiento especial del RE donde es combinado con otras fuentes es posible crear un nuevo producto estadístico más avanzado. En ambos casos se debe aplicar la evaluación de la calidad de entrada.

- 3) Para mejorar el sistema integrado de RE.

En este sentido, se podría mejorar la calidad de las estadísticas generadas a partir de RE y/o los marcos muestrales de encuestas. En este caso se emplea la evaluación de la calidad del proceso de transformación.

- 4) Una combinación de los anteriores.

Calidad de entrada

Puesto que los elementos de entrada provienen de un proceso de recolección de datos que está fuera del control del INE, es de suma importancia comprobar la calidad de la fuente de datos administrativos adquirida, teniendo en cuenta el uso previsto de los datos.

Se recomienda aplicar la herramienta HECRA para evaluar la calidad de dimensión *fuentes de datos administrativas*.

A continuación se detallan las dimensiones, atributos e indicadores utilizados por la herramienta HECRA para evaluar la calidad de entrada.

Calidad asociada con la fuente de datos administrativa	
Atributo	Indicador
I.1. Relevancia	3. Demanda de información (C.I.18)
	4. Satisfacción del usuario interno (C.I.19)
I.2. Entorno institucional, Políticas y procedimientos	1. Marco legal (C.II.20)
	2. Obligación de entrega de datos (C.II.21)
	3. Limitaciones debido a normativa sobre la confidencialidad (C.II.22)
	4. Acuerdos de confidencialidad (C.II.23)
	5. Transferencia de datos segura (C.II.24)
	6. Integridad, Confidencialidad y Disponibilidad de la información (C.II.25)
	7. Protección física de los datos (C.II.26)
	8. Políticas de respaldo de los datos (C.II.27)
I.3. Entrega de los datos	1. Costos asociados a la entrega (C.III.28)
	2. Acuerdos de la entrega (C.III.29)
	3. Frecuencia (C.III.30)
	5. Puntualidad (C.III.32)
	6. Riesgo por falta de datos
	7. Método alternativo para falta de información (C.III.34)
	8. Medio de entrega de los datos (C.III.35)
	9. Formato de archivos (C.III.36)
	10. Unidades de la población objetivo y variables (C.III.30)
	11. Selección de datos (C.III.38)
I.4. Procedimientos	1. Captura de datos (C.IV.39)
	2. Control de consistencia (C.IV.40)
	3. Control de cambios (C.IV.41)
	4. Mejora continua (C.IV.42)
I.5. Tratamiento de los datos	1. Controles de la unidad objetivo (C.V.43)
	2. Controles de contenido de las variables (C.V.44)

	3. Controles de valores extremos (C.V.45)
	4. Modificaciones (C.V.46)
	5. Razón de no modificación (C.V.47)
	6. Modificaciones según procedimiento (C.V.48)
	7. Utiliza manejador de BD (C.V.49)
	9. Documentación de la BD (C.V.51)
	10. Integridad de la BD (C.V.52)

Calidad del proceso de transformación

Las fases y sus correspondientes subprocesos y los posibles errores en unidades y variables se muestran en la siguiente matriz.

Fases y subprocesos del modelo GSRBPM	Errores potenciales (unidades)	Errores potenciales (variables)
1. Especificar necesidades 1.1 Identificar necesidades 1.4 Comprobar disponibilidad de datos 2. Diseño	No hay correspondencia en conceptos (sub-cobertura, sobre-cobertura)	Errores de especificación No hay estabilidad conceptual
4. Recolectar datos 4.1 Organizar/preparar recolecta de datos 4.2 Ejecutar recolecta de datos 4.3 Finalizar recolecta de datos	Errores de selección (falta de respuesta por unidad o missings, duplicados, demoras)	Errores de medición
5. Transformar - 6. Integrar 5.3 Depurar e imputar 6.1 Integrar registros (unir o emparejar) 6.2 Derivar nuevas variables y unidades	Errores de unión de registros Errores de derivación de unidades	Errores de procesamiento (errores de clasificación, datos faltantes, errores de consistencia intra e inter fuente, errores implícitos y explícitos de suposición del modelo)

Fuente: adaptado de Brancato, G. y otros (2016). *Guidelines for the quality of statistical processes that use administrative data*. Istat – Italia.

Evaluación de la calidad de la integración de registros.

El resultado de la integración de datos por cualquier método (determinístico, probabilístico, mixto) debe evaluarse tomando en cuenta dos indicadores principales: coincidencias falsas (falsos positivos) y falsas no coincidencias (falsos negativos). Es difícil encontrar herramientas para realizar este tipo de evaluación, por lo cual es necesario analizar los resultados manualmente, al menos en una muestra de casos emparejados.

Otro aspecto que influye en la validez del procedimiento es la asunción de hipótesis no comprobables: por ejemplo, las reglas aplicadas en el enfoque determinístico o las hipótesis subyacentes a la función de verosimilitud en el modelo probabilístico

Evaluación de la calidad de los metadatos y datos

Los procesos de transformación e integración de registros generan información de metadatos y transforman los datos del RA en RE, por lo cual se debe evaluar la calidad de estas dos dimensiones.

A continuación se detallan los atributos e indicadores utilizados para evaluar la calidad asociada a los metadatos y datos.

Calidad asociada a los Metadatos	
Atributo	Indicador
II.1. Documentación de metadatos	1. Documentación de metadatos (evaluación genérica, más allá de su uso previsto) (D.II.53)
II.2. Completitud y claridad	1. Definición de las unidades de la población (D.II.54)
	2. Descripción de variables (D.II.55)
	3. Comunicación de cambios en definiciones/conceptos (D.II.56)
II.3. Uso de claves únicas	1. Claves de identificación (D.III.57)
	2. Comparabilidad de las variables de identificación (D.III.58)
	3. Combinaciones únicas de variables (D.III.59)
II.4. Comparabilidad	1. Comparabilidad de la definición de la unidad de estudio (D.III.60)
	2. Comparabilidad de la definición del resto de las variables (D.III.61)

Calidad asociada a los datos	
Atributo	Indicador
III.1. Controles técnicos	1. Datos legibles (E.I.62)
	2. Redefinir conceptos y metadatos en el caso de tener más de una fuente (si corresponde) (E.I.63)
	3. Correspondencia con Metadatos (E.I.64)
	4. Método de unión de registros (E.I.65)
	5. Verificación de eficacia del método de unión de registros (E.I.66)
III.2. Cobertura	1. Exceso de cobertura (E.II.67)
	2. Errores de clasificación (E.II.68)
III.3. Unión de registros (informáticos) de los archivos de datos	1. Tasa de casos/filas del archivo unidos con el registro estadístico (E.III.69)
III.4. Completitud	1. Tasa de registros sin datos (E.IV.71)
	2. Tasa de variables sin datos (E.IV.72)
III.5. Medición	1. Control externo al proceso (auditoría interna) (E.V.73)
III.6. Claves de identificación	1. Tasa de registros con clave única (E.VI.74)
III.7. Procesamiento	1. Ajustes (datos editados) (E.VII.75)
	2. Imputación (E.VII.76)
III.9. Uso de códigos y clasificadores estándar (nacional o internacional)	1. Uso de clasificadores estándar (E.IX.79)
	2. Verificación de la codificación (tasa de errores de codificación, tasa de registros sin código por tipo de variable) (E.IX.80)
	3. Tasa de errores de codificación (E.IX.81)
	4. Tasa de registros sin código (E.IX.82)
III.10. Actualidad de los datos	1. El 90% o más de las unidades, que pertenecen a la población objetivo, creadas durante el año t han sido registradas antes del fin del año t+1 (E.X.83)
III.11. Registros múltiples	1. Tasa de registros múltiples de la misma unidad (E.XI.84)

III.12. Otros controles	1. Tasa de unidades que contienen valores válidos en las variables clave (E.XII.85)
	2. Los cuadros de salida de la operación estadística validados a través de métodos automáticos (E.XII.86)

Calidad de salida

El enfoque seguido para evaluar la calidad del producto estadístico final se basa en la evaluación de la calidad de la dimensión “producto estadístico final” de la herramienta HECRA.

A continuación se detallan los atributos e indicadores utilizados para evaluar la calidad de salida del producto estadístico final.

Atributo de la calidad	Indicador de la calidad
Comparabilidad	Longitud de las series temporales comparables (principales indicadores)
	Comparabilidad en el tiempo de los microdatos (principales variables, variables secundarias, variables agregadas, otras variables), si corresponde.
Relevancia	Identificación de los usuarios
	Información sobre los usuarios
	Tasa de satisfacción del usuario (encuesta)
	Utilidad (usos previstos del Registro Estadístico)
Coherencia	Coherencia de estadísticas con periodicidades diferentes
	Coherencia de las estadísticas en el mismo ámbito socioeconómico
Disponibilidad y Transparencia	Accesibilidad por Web
	Tasa de completitud de la información de metadatos del producto estadístico
Precisión y Exactitud	Coefficiente de Variación (encuestas por muestreo basadas en el marco del Registro Estadístico)
	Tasa de falta de respuesta por unidad
	Tasa de falta de respuesta por pregunta
	Tasa de imputación de las principales variables
	Tasa de edición de las principales variables
	Tasa de exceso de cobertura
Oportunidad y Puntualidad	Tasa de errores de clasificación
	Puntualidad en el calendario de difusión/publicación del producto estadístico

Atributo de la calidad	Indicador de la calidad
	Lapso entre el final del período de referencia y la fecha de publicación de los resultados
	Actualidad de los datos del Registro Estadístico

Es importante aplicar el modelo de evaluación de la calidad desde el inicio del proceso de implementación, pues posibilitará realizar un monitoreo y evaluación de la eficacia, eficiencia y calidad del sistema, permitiendo hacer mejoras a medida que se avanza en el proyecto.

11. Metodología para la gestión de inventarios nacionales de RA

En capítulos anteriores se ha indicado que es necesario definir los objetivos y priorizar las áreas temáticas a incluir en el sistema para cada una de las etapas del plan de acción. Esta priorización determinará la demanda de información en cada etapa del proyecto.

La demanda de información permitirá establecer las fuentes administrativas, en particular los registros administrativos necesarios para satisfacer estos requerimientos.

Es fundamental, entonces, crear un inventario de las fuentes de información y sus registros administrativos, donde se podrán caracterizar y describir las principales variables de interés estadístico.

En el anexo II se incluye la metodología para la creación, caracterización, documentación y actualización (gestión) de los registros administrativos de cada fuente administrativa, basada en la *Guía para la gestión de inventarios de registros administrativos de entidades del SEN*, elaborada por el consultor Federico Seguí ¹⁶.

Esta herramienta contiene instrumentos de captura de información acerca de los registros administrativos y las fuentes administrativas, guía de llenado de los formatos de captura y guía de gestión de inventarios de registros administrativos.

12. Sistema integrado de registros estadísticos – SIREPI

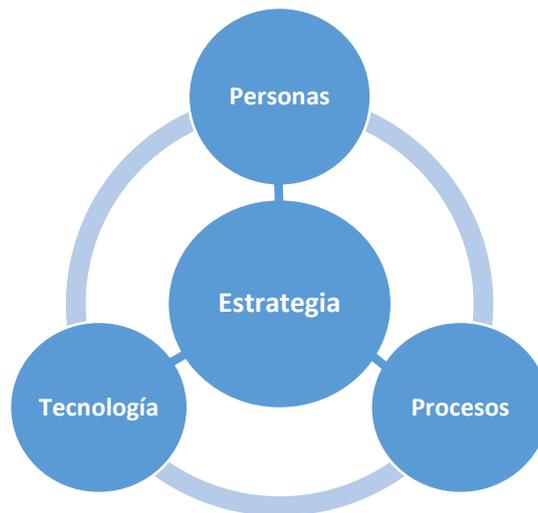
El diseño del SIREPI no forma parte del alcance de la presente consultoría, pero se entiende oportuno mencionar ciertas características básicas que deberían ser consideradas y sentar las bases para el diseño del sistema.

¹⁶ Seguí, Federico (2016). *Guía para la gestión de inventarios de registros administrativos de entidades del SEN*. innovacionestadistica.com

12.1. Características básicas del SIREPI

El SIREPI debería ser diseñado tomando en cuenta un modelo de gestión basado en tres pilares fundamentales: personas, procesos y tecnología, es decir, que el diseño del SIREPI no esté enfocado en la tecnología exclusivamente, sino que además esté alineado con los procesos y éstos con la estrategia.

Figura 24. Modelo de gestión base del SIREPI.



En cuanto a los aspectos tecnológicos, se debería considerar un sistema flexible, parametrizable, lo que comúnmente se llama software generalizado. Debería ser escalable, para que acompañe el crecimiento del proyecto en cuanto a cantidad de RA, volumen de información, capacidad de procesamiento; y multiplataforma (para facilitar la interoperabilidad), lo que permitirá funcionar en diversas arquitecturas tecnológicas, sistemas operativos y hardware.

Se deberían considerar las siguientes alternativas de implementación del sistema, tomando en cuenta las ventajas y desventajas de cada una:

1. Uso de manejadores de bases de datos y paquetes estadísticos:	2. Desarrollo a medida software específico:	3. Desarrollo a medida software generalizado:	4. Sistema de Integración de datos ETL (Extract, Transform and Load) software generalizado:
✓ Rápida y simple implementación	✓ Administración a través del sistema.	✓ Administración a través del sistema.	✓ Administración a través del sistema.
✓ Bajo costo (si no se requieren licencias adicionales)	✓ Sistema integrado.	✓ Orientado a procesos.	✓ Orientado a procesos.
↓ Administración manual, costosa, tediosa.	▪ Orientado a procesos???	✓ Sistema integrado.	✓ Sistema integrado.
↓ No orientado a procesos.	↓ Implementación más compleja.	✓ Flexible, parametrizable.	✓ Flexible, parametrizable.
↓ No sistema integrado.	↓ Mayor costo de desarrollo	✓ Menor costo de mantenimiento	✓ Menor costo de mantenimiento
	↓ Alto costo de mantenimiento/modificaciones	• Implementación menos compleja.	✓ Implementación mas simple.
	↓ No flexible, no parametrizable	↓ Muy alto costo de desarrollo	✓ Bajo costo de desarrollo.
			✓ Corta curva de aprendizaje.

Lo recomendable sería utilizar herramientas de software para la integración de datos denominadas ETL (por su sigla en inglés correspondiente a Extract, Transform and Load, o Extracción, Transformación y Carga) para la recopilación de datos de diferentes fuentes y formatos (extracción), la depuración, estandarización de variables e integración de registros (transformación) y la carga de datos en la base de datos, datawarehouse o datamart (registro estadístico).

Existen varias herramientas ETL en el mercado, dentro de las open source encontramos a una de las más populares como Pentaho Data Integration, y además la suite Pentaho dispone de herramientas de Business Intelligence y Business Analytics para el análisis y explotación de datos.

Las herramientas ETL open source permitirían iniciar en forma gradual, por etapas, sin necesidad de realizar una gran inversión al inicio. Luego, cada INE podría determinar si continúa con la versión “Community” open source o si en cambio invierte en la versión empresarial de pago.

12.2. Bases para el diseño del SIREPI

Se entiende entonces que las herramientas ETL son la opción más adecuada para la automatización de procesos clave de un sistema de las características del SIREPI. Sin embargo, existen una serie de procesos manuales como la realización de convenios o acuerdos que se deben hacer con las entidades para transferir archivos, integrar bases de datos o trabajar con web-services, que deben formalizarse y considerarse como parte del sistema (considerando al sistema en su acepción más amplia: “conjunto de elementos organizados y relacionados que interactúan entre sí para lograr un objetivo”, y no sólo como un sistema informático).

Por tanto se debería trabajar en la estandarización de estos mecanismos de intercambio de información con las entidades y definir protocolos para todos los casos, para facilitar el proceso de **extracción** de datos.

La **transformación** de datos se ha explicado en detalle en el apartado 8.3 Transformación e integración de RA en RE. Las técnicas de machine learning o aprendizaje automático que forman parte de las herramientas de ETL podrían ser una muy buena alternativa para la unión de registros.

Es importante aplicar técnicas de *blocking*, *indexing* o *filtering*, para trabajar con bloques del archivo de menor tamaño y así solucionar el problema de manejar archivos de datos demasiado grandes, y complementarlo con herramientas de *Big data* para manejar grandes volúmenes de información.

El tercer paso o macro proceso es la **carga** de datos en la base de datos central y en data warehouse del SIREPI, si corresponde.

“Un Data Warehouse o Almacén de Datos es un sistema que extrae, depura, ajusta y entrega/publica datos fuente a un almacén de datos dimensionales, para luego soportar e implementar consultas y análisis para la toma de decisiones”¹⁷.

El SIREPI debería tener una base de datos central de producción estadística conformada por las tablas correspondientes a los registros estadísticos que forman parte del sistema y las tablas auxiliares necesarias para el funcionamiento del mismo.

Además, se debería crear un data warehouse a partir de los datos operativos de la base de datos central. Este almacén de datos permitiría explotar al máximo el sistema de registros realizando análisis de datos a través de herramientas de business intelligence.

Es imprescindible implementar mecanismos de protección de datos en las tablas de producción estadística, mediante la desidentificación de los registros estadísticos del sistema (sustituir las variables clave de identificación por claves aleatorias).

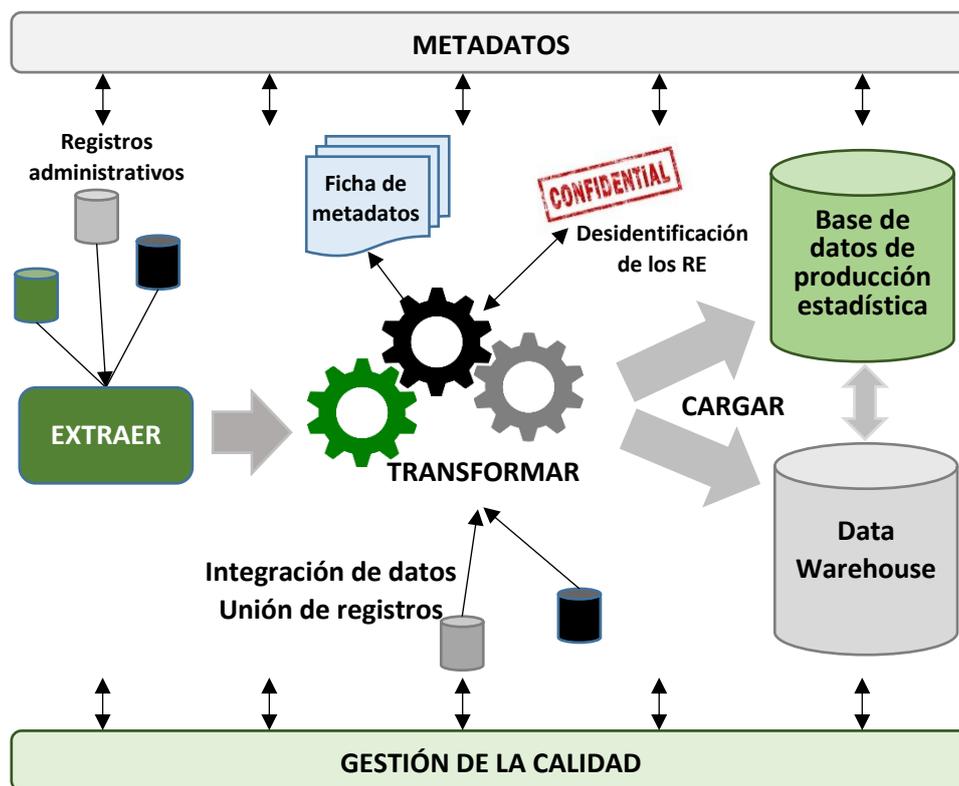
El sistema debe asegurar la trazabilidad de las operaciones realizadas por los usuarios con los datos. Debería contar con un módulo de auditoría que permita consultar el registro o log de auditoría (qué, quién y cuándo se hizo cada operación en el sistema). Asimismo, se debería conservar una historia de los cambios y permitir volver atrás o recuperar estados anteriores.

El SIREPI debería contar con herramientas de gestión de Metadatos y calidad. El módulo de gestión de metadatos debería manejar las fichas de metadatos descritas en el apartado 8.3.2.2 Estandarización de variables. El módulo de evaluación de la calidad debería aplicar las herramientas del modelo de evaluación de la calidad presentado en el capítulo 10.

La siguiente figura presenta la propuesta de diseño básico que tendría el SIREPI.

¹⁷ Kimball, R. Caserta, J (2004). The data warehouse ETL toolkit. Practical techniques for extracting, cleaning, conforming, and delivering data. Wiley. USA.

Figura 25. Diseño básico del SIREPI



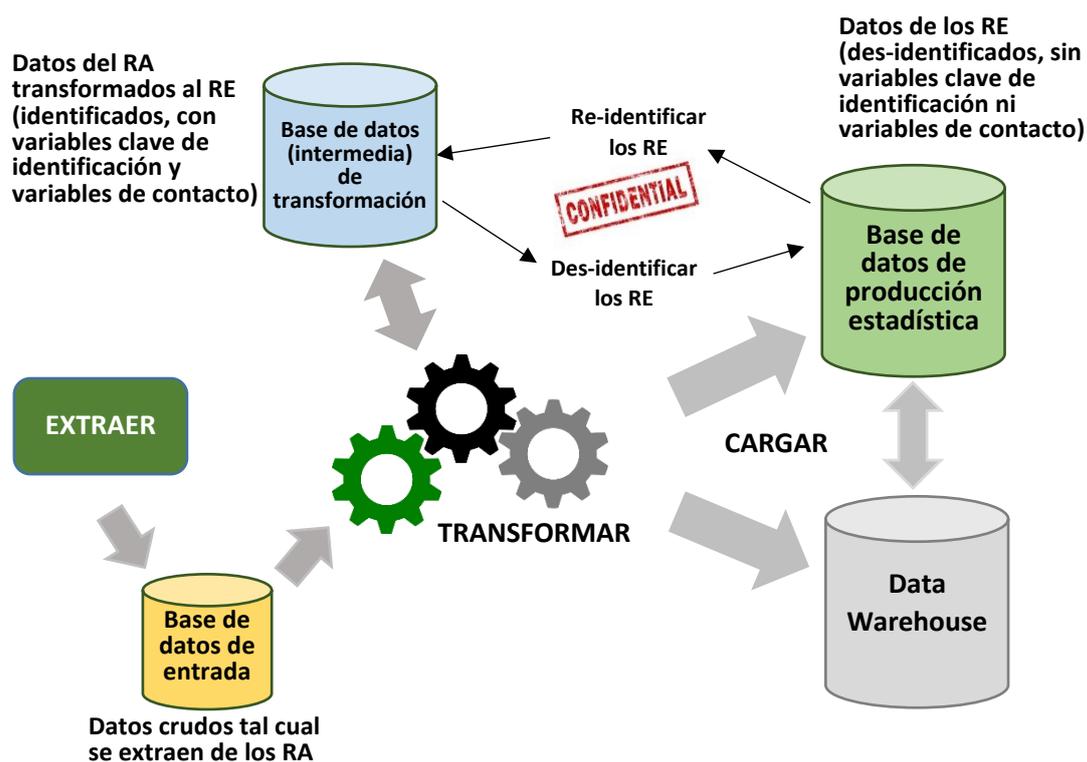
Fuente: elaboración propia del consultor.

Los procesos de gestión de la calidad y gestión de metadatos deben funcionar en paralelo a los procesos misionales de ETL (extracción, transformación y carga) del SIREPI, de forma tal que en cada etapa o actividad se recojan los metadatos correspondientes y se evalúe la calidad a lo largo de todo el proceso.

La aplicación que soporta el SIREPI debería contemplar tres aspectos fundamentales: i) seguir el diseño básico plasmado en la figura anterior, ii) estar alineado con el proceso general descrito por el modelo GSRBPM y iii) permitir diseñar el flujo del proceso adaptado a cada país, partiendo del diseño de procesos planteado en el apartado 6.3.

El sistema debería contar con tres esquemas de bases de datos o grupos de tablas, de acuerdo al esquema de la siguiente figura.

Figura 26. Estructura de esquemas de la base de datos del SIREPI.



Fuente: elaboración propia del consultor.

El sistema extrae los datos provenientes las fuentes administrativas y los almacena en una base de datos de entrada (cilindro naranja) sin ningún procesamiento (datos crudos).

Luego, los datos del registro administrativo almacenados en la base de datos de entrada son transformados a un registro estadístico y son almacenados en una base de datos intermedia (cilindro azul) mientras se realiza la transformación (procesamiento de datos e integración de registros).

Posteriormente, los datos del registro estadístico en la base de datos (intermedia) de transformación se des-identifican para preservar la confidencialidad de la información. En este proceso se eliminan las variables clave de identificación y los identificadores explícitos (nombres, dirección, teléfono, email, coordenadas geográficas, etc.) y se sustituyen por variables clave aleatorias. Finalmente, los datos se cargan des-identificados en la base de datos de producción estadística (cilindro verde) y en el Data Warehouse (cilindro gris).

En el proceso de transformación, cada vez que sea necesario unir o emparejar nuevos registros con registros base o registros estadísticos existentes en la base de datos de producción estadística, se procederá a re-identificar dichos registros (se le vuelven a incorporar las variables clave de identificación y los identificadores explícitos) y se copian estas tablas a la base de datos de

transformación para que sea posible la unión de registros. Las variables con datos confidenciales serán protegidas para evitar que los usuarios que realizan la transformación de registros puedan acceder a dichos datos y minimizar el riesgo de divulgar información confidencial.

13. Glosario

Aprovechamiento o explotación de Registros Administrativos (con fines estadísticos). “método para generar datos estadísticos, mediante el uso de los sistemas de registro de hechos o sucesos individuales que realizan las dependencias y organismos públicos como parte de su función”. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). DGE. DGAIN. DN. Glosario sobre la Generación de Estadística Básica. Marzo de 2006.*

Archivo de microdatos. Es un arreglo matricial de datos individuales (personas, viviendas, establecimientos, u otras unidades de observación) en medios computacionales, donde las filas representan a cada unidad o individuo y las columnas son las variables que contienen información (datos) sobre las características de la población.

Calidad. Existen varias definiciones de calidad universalmente aceptadas. Algunas de ellas son:

“Propiedad o conjunto de propiedades inherentes a una cosa que permiten apreciarla como igual, mejor o peor que las restantes de su especie”. *Fuente: Real Academia de la lengua española.*

“Grado en el que un conjunto de características inherentes cumple con los requisitos”. *Fuente: Norma ISO 9000.*

“Calidad es adecuación al uso del cliente”. *Fuente: Joseph Juran.*

Captación. Serie de actividades para obtener los datos de cada unidad objeto de registro, siguiendo las estrategias determinadas en los programas y procedimientos de trabajo. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Captación en Registros Administrativos. Serie: Lineamientos para la generación de estadística básica. 2010.*

Captura de datos. Procedimiento para transformar la información del cuestionario en un archivo electrónico de datos. *Fuente: ONU. (2001); Departamento de Asuntos Económicos y Sociales. División de Estadística. Manual sobre Gestión de Censos de Población y Habitación. Estudios de Métodos. Serie F. No. 83. Naciones Unidas, Nueva York. p: 157*

Codificación. Procedimiento para asignar identificadores numéricos o alfanuméricos a conceptos en un orden establecido. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Glosario sobre la Generación de Estadística Básica. Marzo de 2006.*

Criterios de validación. Conjunto de reglas de naturaleza conceptual y estadística, que sirven de base para la identificación y solución de los problemas que se presentan en los datos estadísticos. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Diseño conceptual. Serie: Documentos técnicos para la generación de estadística. 2006.*

Dato administrativo. Valor de una variable correspondiente a una acción, hecho o evento que forma parte de un Registro Administrativo.

Dato estadístico. Valor cuantitativo de un conjunto específico respecto a una variable, con referencia de tiempo y de espacio. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Glosario sobre la Generación de Estadística Básica. Marzo de 2006.*

Edición de datos: procedimiento para detectar y corregir datos que están en blanco o no cumplen con ciertas reglas de consistencia de la información. Según la definición de la Organización para la Cooperación y el Desarrollo Económico (OCDE), la edición de datos es la actividad que tiene por objetivo detectar y corregir errores (inconsistencias lógicas) en los datos. Las técnicas de edición se refieren a ciertos procedimientos y procesos utilizados para detectar y manejar los errores en los datos.

Estadísticas basadas en registros administrativos: elaboración de productos estadísticos (ver definición) a partir del procesamiento y la transformación de registros administrativos en registros estadísticos.

Fuente de datos administrativa o fuente administrativa: es definida por la OCDE como: “la unidad de organización responsable de implementar una regulación administrativa (o grupo de regulaciones), cuyo registro correspondiente de unidades y transacciones se ven como fuente de datos estadísticos”. Dependencia u organismo público, privado o mixto responsable del Registro Administrativo y sus oficinas donde se llevan a cabo los procesos de captación y mantenimiento de los datos del registro. El Registro Administrativo es el resultado de la acción de registrar los datos de determinado evento administrativo; mientras que la fuente administrativa se refiere a la organización responsable del RA, los procesos y entorno institucional que involucran al mismo. Ver definición de Registro Administrativo.

Fuente informante: entidad o persona que informa sobre la acción, evento o hecho objeto de registro.

HECRA o Herramienta. Herramienta para la Evaluación de la Calidad de Registros Administrativos a ser usados con fines estadísticos, elaborada por el consultor internacional Federico Segui para el Banco Mundial.

Imputación de datos: procedimiento estadístico para asignar valores a un dato faltante o en sustitución de valores de respuesta no válidos o inconsistentes. El objetivo es reemplazar los valores faltantes, erróneos o inconsistentes utilizando variables auxiliares, por medio de procedimientos estadísticos estandarizados.

Indicador. Es un parámetro o un valor derivado de parámetros, que proporciona información y/o describe el estado de determinado fenómeno, y tiene un significado que se extiende más allá de aquel directamente asociado a cualquier valor paramétrico dado.

“Herramientas para clarificar y definir, de forma más precisa, objetivos e impactos (...) son medidas verificables de cambio o resultado (...) diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso (...) con respecto a metas establecidas, facilitan el reparto de insumos, produciendo (...) productos y alcanzando objetivos”. *Fuente: Organización de las Naciones Unidas*

(ONU). *Integrated and coordinated implementation and follow-up of major United Nations conferences and summits*. Nueva York, Estados Unidos de América, 10 y 11 de mayo de 1999, p. 18. Consultado en internet en la página www.un.org/documents/ecosoc/docs/1999/e1999-11. 17 de julio de 2011.

Una de las definiciones más utilizadas por diferentes organismos y autores es la que Bauer dio en 1966: "Los indicadores sociales (...) son estadísticas, serie estadística o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto".
Fuente: Horn, Robert V. Statistical indicators for the economic and social sciences. Cambridge, University Press, Hong Kong, 1993, p. 147.

Indicador estadístico. Es un elemento de datos que representa datos estadísticos para determinado período, lugar y otras características. *Fuente: Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000.*

Instrumento de captación. Formato que se utiliza para el registro de los datos a nivel de cada unidad objeto de registro. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Glosario sobre la generación de estadística básica. Marzo de 2006.*

Marco legal. Conjunto de leyes, reglamentos, políticas y normas que fundamentan jurídicamente los registros que las dependencias y organismos de la administración pública realizan como parte de su función. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Glosario sobre la generación de estadística básica. Marzo de 2006.*

Medio del formato en que se hace el registro. Tipo de instrumento físico utilizado para contener los formatos. Puede ser de dos tipos: impreso y electrónico. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Captación en Registros Administrativos. Serie: Documentos técnicos para la generación de estadística básica. 2010.*

Metadatos estadísticos. Información acerca de los datos estadísticos. Los metadatos comprenden datos y otra documentación que describe objetos de una manera formalizada. *Fuente: UNECE. "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000.* Los metadatos brindan información sobre los datos estadístico y los procesos de producción y uso de los mismos. *Fuente: UNECE, "Guidelines for the Modelling of Statistical Data and Metadata", 1995.*

Microdato. Cada uno de los datos referentes a cada una de las unidades de observación obtenidos en un proyecto de estadística. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Procesamiento de la información. Serie: Lineamientos para la generación de estadística básica. 2006.*

Operación estadística: conjunto de procedimientos y actividades llevados a cabo para elaborar uno o varios productos estadísticos.

Población de interés o estudio: son todos los casos o unidades que forman parte del Registro Estadístico que cumplen con un conjunto de características particulares, es decir, ciertas variables tienen determinados valores en común. La población de interés o estudio puede o no coincidir con la totalidad de unidades que componen el Registro Estadístico. Es posible, entonces, definir a la población de interés como un subconjunto del Registro Estadístico, esto va a depender del uso con fines estadísticos que se quiera hacer del mismo.

Población objetivo: ver Población de interés o estudio.

Procesamiento de datos. Serie de actividades para preparar los archivos de datos, asegurándose que sean congruentes y ordenados para su aprovechamiento. *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Glosario sobre la Generación de Estadística Básica. Marzo de 2006.*

Procedimiento documentado. Documento que describe la secuencia de actividades para ejecutar efectivamente un proceso. Además, significa que el procedimiento se ha establecido, documentado, implementado y mantenido.

Proceso: conjunto de actividades mutuamente relacionadas o que interactúan, las cuales transforman elementos de entrada en resultados. *Fuente: ISO 9000 - Sistemas de gestión de la calidad - Conceptos y vocabulario. 2000.*

Producto estadístico: uno o varios indicadores, cuadros estadísticos y/o archivos de microdatos preparados para ser aprovechados con fines estadísticos, ya sea para la toma de decisiones, definición de políticas públicas o análisis de series históricas.

Registro Administrativo (RA): "todo registro resultante de necesidades fiscales, tributarias u otras, creado con la finalidad de viabilizar la administración de los programas de gobierno o para fiscalizar el cumplimiento de obligaciones legales de la sociedad". *Fuente: CEPAL, II CEA, 2003: 10.*

"Serie de datos sobre un tipo de sujeto, acción, hecho o evento, obtenidos mediante un proceso de captación, con base en un formato específico ya sea impreso en papel o en medios computacionales, y que realiza una institución pública, bajo un marco de funciones y facultades formalmente establecidas en instrumentos jurídicos, reglamentarios o programáticos". *Fuente: Instituto Nacional de Estadística y Geografía de México (INEGI). Proceso estándar para el aprovechamiento de Registros Administrativos. Serie: Documentos técnicos para la generación de estadística básica. 2010.* Esta última definición será la utilizada en el presente documento.

Registro con fines estadísticos o Registro Estadístico (RE): es el registro consolidado de datos estandarizados y procesados provenientes de uno o más Registros Administrativos, que originalmente no (necesariamente) fueron captados con fines estadísticos, pertenecientes a una o más fuentes de datos administrativos, pero que han sido adaptados para su uso estadístico.

Regulación administrativa. Son los trámites y las formalidades administrativas con que los gobiernos recogen información e intervienen en decisiones económicas individuales. *Fuente: OCDE. Glosario de Términos Estadísticos.*

Responsable del Registro Administrativo. Unidad o cargo responsable del mantenimiento de los datos del Registro Administrativo. Ver también: Fuente de datos administrativos.

Seguridad de la información - Confidencialidad, Integridad y Disponibilidad. La seguridad de la información tiene por objetivo la preservación de la confidencialidad, integridad y disponibilidad de la información. La confidencialidad se refiere a la información que no se revela ni se encuentra a disposición de individuos, organizaciones o procesos no autorizados. La integridad es la propiedad de salvaguardar la exactitud y completitud de la información. La disponibilidad está referida a la información que puede ser accesible y utilizable a pedido de un agente autorizado. *Fuente: ISO 27001:2005 – Sistemas de Gestión de Seguridad de la Información. Requerimientos.*

Sistema integrado de registros estadísticos: Conjunto ordenado de principios y procedimientos que establecen la interrelación e interacción de los registros estadísticos que forman parte del sistema.

Transferencia de datos administrativos: es el envío de los datos del Registro Administrativo al usuario primario, por parte de la fuente administrativa.

Unidad de la población: caso o unidad que forma parte de la población de interés del Registro Administrativo. Las unidades de la población cumplen con un conjunto de características particulares, es decir, ciertas variables tienen determinados valores en común.

Usuario final: es quien recibe finalmente el producto estadístico. Se trata de quien consume la información estadística contenida en la publicación final sobre el fenómeno de estudio basado en el Registro Administrativo. En algunos casos podrá tratarse de un usuario de la misma institución responsable del Registro Administrativo, y en otros casos podrán ser usuarios externos a ésta.

Usuario primario del Registro Administrativo o productor estadístico: es el usuario directo -en ocasiones interno- de los datos administrativos. En algunos casos puede tratarse del estadístico que trabaja en la Oficina Nacional de Estadística o Departamento de estadística de la institución responsable del Registro Administrativo, cuyo objetivo es producir indicadores sobre un fenómeno en particular, es decir, quien explota el registro con fines estadísticos.

Validación de datos. Actividades de control realizadas en los campos de un registro en particular de un archivo de datos. Esto incluye la comprobación de cada campo de cada registro para determinar si contiene un dato válido (criterios de validación) y la comprobación de que los datos de todos los campos son coherentes entre sí.

Variable. Una variable es una característica de una unidad observada que puede

asumir más de uno de un conjunto de valores a los que se puede asignar una medida numérica o una categoría de una clasificación (por ejemplo, ingreso, edad, peso, etc., y "ocupación", "Industria", "enfermedad", etc.). *Fuente: OCDE. Glosario de Términos Estadísticos.*

Variable estadística. Una variable es un atributo medible de un objeto o unidad estadística. Una variable estadística está definida por el tipo de objeto que presenta la característica (por ejemplo, ingreso para personas e ingreso para hogares son dos variables distintas), por el método de medición y la escala aplicados, y por el momento o período a los que refiere la medición.

14. Bibliografía

Baxter, R. Gu, L. Vickers, D. Rainsford, C (2016). *Record Linkage: Current Practice and Future Directions*. CSIRO Mathematical and Information Sciences. CMIS Technical Report No. 03/83. Canberra, Australia.

Berning, M. y otros (2013). *Data Quality Assessment Tool for Administrative Data*. U.S. Census Bureau. USA.

Biemer P., Lyberg L. (2003). *Introduction to survey quality*. Wiley, New York.

Blomqvist, Klas and others (2011). *A strategy to improve the register system to store, share and access data and its connections to a generic statistical information model (GSIM)*. Invited paper. Work Session on Statistical Data Editing of the Conference of European Statisticians – UNECE. Ljubljana, Slovenia, 9-11 May 2011.

Booch, Grady (2006). *The Accidental Architecture*. IEEE Software. Mayo-Junio 2006.

<https://pdfs.semanticscholar.org/7fdc/6cde5c79f046e4a33e4fb36210abb041b0c6.pdf>

Brancato, G. y otros (2016). *Guidelines for the quality of statistical processes that use administrative data*. Istat. Italia.

Bycroft, C (2010). *A register-based census: what is the potential for New Zealand?* Wellington. Statistics New Zealand.

Elfeky, M. y otros (2003). *Record Linkage: A Machine Learning Approach, A toolbox, and a Digital Government Web Service*. Computer Science Technical Reports. Paper 1573. Purdue University. USA.
<http://docs.lib.purdue.edu/cstech/1573>

Fellegi I.P., Sunter A.B. (1969) *A theory for record linkage*. Journal of the American Statistical Association 64, 1183-1210. USA.

Fingar, P. Smith, H (2003). *Business Process Management. The third wave*. Meghan-Kiffer Press. USA.

Kimball, R. Caserta, J (2004). *The data warehouse ETL toolkit. Practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley. USA.

Laitila, T., Wallgren, A. Wallgren B. (2011). *Quality Assessment of Administrative Data*. Research and Development – Methodology reports from Statistics Sweden 2011:2. Suecia.

Medina, Alejandro. Segui, Federico (2013). *¿Cómo mejorar el desempeño y crear valor público en las oficinas y sistemas estadísticos nacionales en América Latina y el Caribe?* CreateSpace. USA. innovacionestadistica.com

Segui Stagno, Federico (2009). *Implementing ISO 20252:2006 “Market, opinion and social research” in a statistical office*. Special Topic Paper. Actas del 57º Congreso Mundial de Estadística, ISI 2009, Durban - Sudáfrica).

Segui Stagno, Federico (2011). *Certifying the Quality & Information Security Management Systems of the NSO according to the international standards ISO 9001, ISO 20252 and ISO 27001*. Special Topic Paper. Actas del 59º Congreso Mundial de Estadística, ISI 2011, Dublín-Irlanda.

Segui Stagno, Federico (2012). *Guía de la herramienta para la evaluación de la calidad de Registros Administrativos (HECRA) a ser usados con fines estadísticos*. Banco Mundial.

Segui Stagno, Federico y otros (2012). *Quality improvement of administrative registers statistically exploited to generate the indicator-based decision-making system in the State of Yucatan, Mexico*. Actas de la Conferencia sobre Estadísticas Oficiales IAOS 2012. Kiev, Ucrania.

Segui Stagno, Federico (2016). *Guía para la gestión de inventarios de registros administrativos de entidades del SEN*. innovacionestadistica.com

UNECE (2007). *Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics*. Naciones Unidas. Nueva York y Ginebra, 2007.

UNECE (2011). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. Naciones Unidas. Nueva York y Ginebra, 2011.

UNECE (2013a). *The Generic Statistical Business Process Model GSBPM v5.0*
<http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

UNECE (2013b). *Generic Statistical Information Model GSIM v1.1*.
<http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model>

UNECE (2015). *Common Statistical Production Architecture*. UNECE.
<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture>

Wallgren, A. Wallgren, B. (2012). *Estadísticas basadas en registros. Aprovechamiento estadístico de los registros administrativos*. INEGI.

Wallgren, A. Wallgren, B. (2016). *Hacia un sistema estadístico integrado basado en registros*.

Zhang L.C. (2012). *Topics of statistical theory for register-based statistics and data integration*. Statistica Neerlandica (2012) Vol 66, no.1, pp. 41-63.

Anexo I – Síntesis del diagnóstico subregional

INTRODUCCIÓN

En este informe de síntesis de diagnóstico subregional se busca identificar los factores críticos que impactan en la producción, acceso y uso de los registros administrativos en la subregión y los mínimos comunes a los cuatro países para la conformación de un sistema integrado de registros estadísticos de población e inmuebles.

Factores como la participación activa de los responsables de la producción estadística, compartiendo su conocimiento técnico para alcanzar los objetivos del proyecto, apoyo de la alta dirección, marco normativo, entorno institucional, recursos humanos capacitados y tecnología disponible son elementos clave para poder avanzar en un proyecto de integración de registros administrativos con fines estadísticos. Estos factores están presentes de manera diferencial en los cuatro países, tal como se describirá más adelante.

El trabajo inicial con la aplicación de los instrumentos de evaluación de la calidad desarrollados específicamente para este proyecto permitió contar con una primera aproximación al estado de situación en los países sobre el uso de registros administrativos con fines estadísticos, plasmado en los informes de diagnósticos nacionales.

Al igual que en cualquier investigación, existen errores debido a diversos factores como son los instrumentos de diagnóstico utilizados, la aplicación incorrecta de los instrumentos o interpretación errónea de las preguntas, la falta de conocimiento cabal del ámbito de investigación (registros administrativos evaluados tanto desde el punto de vista administrativo como su uso potencial para producir estadísticas) y falta de tiempo para realizar el diagnóstico de acuerdo a los objetivos y resultados esperados.

En el **Anexo I.1** se encuentran las **observaciones y recomendaciones para la mejora de los instrumentos de evaluación utilizados en los diagnósticos nacionales**.

En el siguiente apartado se describe la metodología utilizada para la elaboración de esta síntesis.

Las visitas realizadas a los países en el marco de esta consultoría resultaron fundamentales y sirvieron para obtener información complementaria a los diagnósticos nacionales, así como para ratificar o rectificar los hallazgos de éstos.

También fue de gran utilidad la Ficha de requerimientos de información a cada país (Anexo I.2) que facilitó la recogida de información adicional previo a las visitas. Un aspecto relevante que se ha detectado a través de estas fichas, entre otros, es que los INE no disponen de información precisa y detallada sobre la calidad de los datos de los registros administrativos.

Es así que se ha llegado a este documento que pretende ahondar en ciertos aspectos no cubiertos o parcialmente evaluados por los diagnósticos nacionales y brindar una síntesis de la situación de la subregión en cuanto a la calidad de los registros administrativos, potencial para su uso en la producción estadística y la capacidad institucional para la operación de registros administrativos.

METODOLOGÍA

La síntesis del diagnóstico subregional ha sido elaborada en dos etapas. En la primera etapa se realizó la síntesis preliminar del diagnóstico subregional exclusivamente sobre la base de los diagnósticos nacionales elaborados por los consultores contratados o funcionarios de cada INE y otros documentos proporcionados por la coordinación del proyecto. La segunda parte implicó la visita a los cuatro países para complementar la información de los diagnósticos nacionales, para luego actualizar la versión final de la síntesis del diagnóstico subregional.

Cabe señalar que los posibles errores u omisiones en la presente síntesis del diagnóstico subregional inducidos por errores u omisiones de los mencionados informes son de entera responsabilidad de los autores de dichos informes, sus colaboradores y/o fuentes de información.

La metodología utilizada para elaborar la síntesis del diagnóstico subregional consta de los siguientes elementos y actividades:

a) Análisis de los documentos e informes preparados durante las etapas anteriores del proyecto:

- Cuestionarios CECRA y CECAP.
- Guías de llenado de los cuestionarios CECRA y CECAP.
- Informe final de la consultora Josefina Calva de febrero de 2015.
- Actas de reuniones y memorias de los talleres.
- Informes de diagnóstico nacional de los registros administrativos (RA) de población e inmuebles.
- Matriz integrada de evaluación de la calidad.
- Matriz integrada de evaluación de la capacidad institucional.
- Matriz integrada de mecanismos actuales de coordinación.
- Matriz integrada de condiciones para la conformación del SIREPI.

b) Reuniones virtuales por medio de video y tele conferencias con los INE de cada país.

Se han organizado reuniones virtuales por medio de videoconferencias con los técnicos de los INE de cada país y el consultor a los efectos de evacuar dudas y consultas sobre la situación actual de cada país y los informes de diagnóstico nacional.

Estas reuniones virtuales por videoconferencia o conversaciones por teleconferencia han sido utilizadas a lo largo de todo el proyecto como mecanismo de comunicación permanente.

c) Intercambio de información por medios electrónicos.

Se utilizaron medios electrónicos de intercambio de información como el correo electrónico o espacios en la nube con acceso restringido mediante claves, para facilitar el envío de información entre los técnicos de los INE y el consultor.

- d) La experiencia del consultor en la región de Latinoamérica y el Caribe en evaluaciones análogas ha sido de gran utilidad para elaborar la síntesis del diagnóstico subregional.
- e) Evaluación y propuestas de mejora de los instrumentos y criterios utilizados para realizar las evaluaciones de calidad y capacidad institucional.
- f) Elaboración de una matriz comparativa preliminar con resultados que permitan identificar los aspectos comunes y particulares de cada uno de los cuatro países sobre el uso estadístico de los RA de población e inmuebles.
- g) Identificación de los vacíos de información para establecer los mínimos comunes entre los países (RA comunes de población e inmuebles. Incluyendo variables, desagregaciones geográficas e identificadores comunes).
- h) Elaboración de la ficha de requerimientos de información a cada país de acuerdo a los hallazgos encontrados (ver Anexo I.2).
- i) Preparación de las visitas a los países para verificar la información proporcionada con evidencias objetivas, la articulación con los responsables de los registros y los aspectos comunes propuestos en la revisión de los diagnósticos y la matriz preliminar.
- j) Visitas a los cuatro países para mantener reuniones con los técnicos de los INE y entidades responsables de los RA.

Se han realizado visitas a cada país para recabar información adicional y validar la información proporcionada por cada país. Esta información ha sido de suma importancia para actualizar la síntesis preliminar del diagnóstico subregional.

Estas reuniones con los técnicos de los INE permitieron identificar la experiencia y disposición de los funcionarios tanto para con el proyecto como para desarrollar sinergias entre los puntos focales del proyecto y los técnicos de las áreas encargadas de la producción estadística.

- k) Actualización del informe final de la síntesis del diagnóstico subregional.

SÍNTESIS DEL DIAGNÓSTICO SUBREGIONAL

El presente apartado está orientado al análisis, interpretación y consolidación de los resultados de los diagnósticos nacionales, capacidades institucionales e identificación de los factores comunes y no comunes en los Sistemas Estadísticos de los países de la CAN.

Las siguientes matrices de síntesis subregional han sido elaboradas a partir de la matriz integrada de evaluación de la calidad de los RA creada durante el taller regional del mes de mayo de 2016¹⁸ y corresponde al promedio de las calificaciones de los RA evaluados en cada país.

Esta evaluación se compone por una parte a los RA evaluados en los diagnósticos nacionales, y otra por evaluaciones incorporadas con posterioridad a éstos por parte de cada INE con motivo del taller regional. En este último caso, los técnicos de los INE realizaron las evaluaciones que no necesariamente son comparables con los diagnósticos elaborados por los consultores nacionales, pues no se aplicaron los instrumentos desarrollados para dichos diagnósticos.

Tabla 1. Matriz de síntesis subregional de evaluación de la calidad de los RA.

País	Accesibilidad	Credibilidad	Coherencia	Oportunidad	Pertinencia	Precisión
Bolivia	1,8	2,4	2,7	2	2,6	2,5
Perú	2,2	2,4	2,8	2,4	2,9	2,9
Ecuador	3	2,6	2,6	2,6	2,7	2,1
Colombia	2,9	2,3	3,0	2,8	1,5	2,4

Fuente: elaborada a partir de los documentos del taller regional 2016.

Tabla 2. Matriz de síntesis subregional de evaluación de la calidad de los RA, según tipo.

País	Tipo de RA	Accesibilidad	Credibilidad	Coherencia	Oportunidad	Pertinencia	Precisión
Bolivia	Pob	2,3	2,3	2,6	2	2,6	2,5
	Inm	0,9	2,5	2,9	2	2,6	2,4
Perú	Pob	1,8	2,2	2,2	2,4	2,64	2,5
	Inm	2,2	2,4	2,8	2,4	2,94	2,9
Ecuador	Pob	3	2,6	2,6	2,6	2,7	2,1
Colombia	Pob	3	2,1	3	2,3	1	2,5
	Inm	2,7	2,3	3	2,5	2,5	2,3

Fuente: elaborada a partir de los documentos del taller regional 2016.

¹⁸ Taller Síntesis de Diagnóstico Regional. 24 al 27 de mayo de 2016, Bogotá – Colombia.

Tabla 3. Matriz integrada de evaluación de la capacidad institucional.

Dimensión	BOLIVIA			ECUADOR			PERÚ			COLOMBIA		
	Pob.	Inm.	Promedio	Pob.	Inm.	Promedio	Pob.	Inm.	Promedio	Pob.	Inm.	Promedio
Recursos humanos	3,0	1,8	2,4	1,5	1,8	1,6	1,0	1,0	1,0	2,3	2,1	2,2
Recursos Institucionales	2,0	1,5	1,8	2,38	2,0	2,2	0,1	0,8	0,4	1,5	1,9	1,7
Validez metodológica	1,8	2,0	1,9	2,5	2,5	2,5	1,6	1,2	1,4	1,4	1,6	1,5
Control de calidad	1,0	1,3	1,2	1,67	1,67	1,7	0,7	0,3	0,5	1,5	1,5	1,5

Fuente: documentos del taller regional 2016.

Es oportuno hacer ciertas precisiones como que en el caso de Colombia se han evaluado solo tres RA, lo cual es insuficiente como para extrapolar los resultados al resto de los RA que podrían formar parte de un sistema integrado o llegar a conclusiones sólidas, o el caso de Ecuador que sólo se han considerado los registros de población en la matriz integrada de evaluación de la calidad. Asimismo, como se ha indicado en las observaciones sobre los instrumentos de evaluación utilizados, existen varios factores que podrían afectar los resultados de los indicadores de la calidad del estudio y por lo tanto esta matriz de resumen podría no contar con una adecuada precisión. Por lo cual se recomienda analizar estos resultados con precaución.

Sin embargo, hay ciertas dimensiones de la calidad que parecen estar en consonancia con las percepciones y evaluaciones cualitativas de los consultores y autores de los informes. Así como también, van en la línea de las expresiones e información transmitida por los técnicos de cada INE (durante las reuniones virtuales y presenciales realizadas) acerca de la situación actual de cada país en relación al uso de registros administrativos con fines estadísticos, sobre todo en lo que respecta a la accesibilidad de los datos.

Existen marcadas diferencias entre los países sobre el grado de avance en cuanto al uso estadístico de los RA que de alguna manera se refleja en el indicador de accesibilidad (Tabla 1). Aunque si se compara por tipo de RA de cada país en cuanto al uso estadístico y accesibilidad de los datos de cada uno, las diferencias probablemente serán más notorias.

En el caso del DANE de Colombia y el INEC de Ecuador se ha podido detectar avances sobre el conocimiento y priorización de los RA con respecto al momento en que se hicieron los diagnósticos nacionales y el taller regional del mes de mayo de 2016, pero parecen estar más asociados al interés por implementar los ejercicios piloto en cada país (ejercicio piloto de San Andrés y el censo piloto de Jamundí en el caso de Colombia y el censo piloto de Galápagos en el caso de Ecuador) que por cumplir con los objetivos del proyecto. El INE de Bolivia y el INEI de Perú están en proceso de caracterización y priorización de los RA.

En los cuatro países se encontró que los RA de población e inmuebles presentan contrastes importantes en cuanto a su uso estadístico. En general, los RA de población son más utilizados con fines estadísticos que los RA de inmuebles, como lo muestra la

Matriz de síntesis subregional de evaluación de la calidad de los RA según tipo (Tabla 2). Los INE tienen un mayor acceso a los RA de población que a los de inmuebles. Históricamente, las estadísticas vitales han sido elaboradas a partir de los RA de nacimientos y defunciones, y luego paso a paso se le han ido incorporando otros RA de población.

En general en los RA de población se utiliza una clave estandarizada como la cédula de ciudadanía, DNI o cédula de identidad, pero en el caso de Bolivia no todas las entidades la utilizan como clave de identificación de las personas (algunas entidades usan un código propio y la cédula de identidad es un dato más del registro con lo cual su calidad y completitud no están aseguradas). Las entidades han comenzado a trabajar en mejorar sus sistemas de registros incorporando la cédula de identidad como clave única de identificación de las personas.

Los RA de inmuebles presentan mayores dificultades a la hora de unir los diferentes archivos o bases de datos debido a que no siempre utilizan las mismas claves de identificación de casos.

No existe una clave de identificación de la vivienda asociada a las personas, por lo cual se debe utilizar el domicilio que en la mayoría de los casos no está estandarizado.

En las visitas a los cuatro países se pudo confirmar la presunción planteada por el consultor durante las reuniones virtuales, acerca de que el registro de servicios públicos (en especial el registro de energía eléctrica), en general es uno de los que cuenta con mayor cobertura (a pesar de las conexiones irregulares) dentro de los registros de inmuebles (incluso cuenta con información de coordenadas geográficas de ubicación de los medidores de consumo). Este registro debería integrarse con otros registros de inmuebles para mejorar su cobertura temática, complementándolo con variables provenientes de otras fuentes.

Los primeros pasos serían crear un marco o registro de domicilios estandarizados del país y que todas las fuentes de RA de inmuebles lo utilicen como clasificador estandarizado o nomenclátor de domicilios. Colombia y Ecuador ya han iniciado este camino.

No hay un registro de residencia consolidado y actualizado, como sí existe en los países nórdicos, que permita vincular a los registros de población con inmuebles y mantener actualizada la información sobre los cambios de residencia.

En Perú la entidad responsable del registro civil RENIEC ha iniciado la creación del registro de residencia, pero ha tenido dificultades para mantener actualizada esta información.

En el caso de Ecuador se tiene previsto iniciar con este registro a partir de 2017 y el INEC está apostando mucho a este registro sobre todo para realizar sus censos basados en registros administrativos a partir de 2020.

Cabe señalar que el costo del censo basado en registros administrativos en los países nórdicos es de alrededor de 10% o menos que un censo tradicional, pero sin considerar el costo del mantenimiento de los sistemas administrativos.

Estos países han tenido que establecer la siguiente infraestructura, que les ha llevado entre 20 y 30 años crearla:

- Registro nacional de población (residencia).

- Registro nacional de direcciones (y registro de inmuebles en algún caso).
- Vinculación precisa entre los sistemas administrativos (un único identificador personal).
- Notificación obligatoria de los cambios de domicilio.

Crear y mantener registros administrativos es muy costoso, lo cual es un buen argumento para hacer un mayor uso de los datos administrativos existentes. Sin embargo, esta no es una buena razón para reemplazar los costos de un censo tradicional por los costos de una nueva infraestructura administrativa, a menos que también se deba cumplir un fuerte propósito administrativo.

Los países que tienen un registro nacional de residencia han visto mayores beneficios para la eficiencia y administración del gobierno, y no tan solo para la generación de estadísticas basadas en dichos registros.

En este punto surge la misma reflexión planteada por otros países como Nueva Zelanda¹⁹ y Reino Unido al momento de planificar sus censos basados en registros administrativos: si no se cuenta con un registro de residencia (o registro nacional de población) o no se tiene acceso a éste se debería empezar con lo que se dispone para luego integrar gradualmente otros registros, y en paralelo se comenzaría a crear el registro de residencia. De esta forma no se estaría apostando exclusivamente a dicho registro.

La iniciativa de crear un registro nacional de población (residencia) debería surgir de todo el gobierno y seguramente sea un proyecto a largo plazo. Implicaría definir y acordar los objetivos, coordinación y financiamiento del proyecto; destinar tiempo suficiente para el debate público, modificar la legislación vigente, definir los procesos misionales (inter-institucionales), implementar los sistemas de información. Tal vez lleve dos o tres censos tradicionales lograr un registro administrativo de población (residencia) completamente funcional.

Otro enfoque considerado por la oficina de estadística de Nueva Zelanda es crear un registro estadístico de población integrando registros administrativos existentes que ya tengan una alta cobertura de la población (salud, educación, impuestos, registro electoral). Han estado avanzando en este enfoque, a pesar de que tal vez no es una alternativa factible para realizar un censo. Tienen claro que no proporcionará un reemplazo para el censo en el corto y mediano plazo.

En Nueva Zelanda entonces, están evaluando estas dos alternativas:

- Crear un registro nacional de población (residencia) donde la función administrativa está estrechamente alineada con la necesidad del censo de contar la población. Este enfoque es la base de ejemplos internacionales exitosos. Sin embargo, en el contexto de Nueva Zelandia, habría que identificar otros beneficios para el gobierno y considerar la privacidad y las cuestiones culturales. Pero no será sino hasta dentro de 20 o 30 años que esperan tener este registro.
- Alternativamente, la vinculación de múltiples fuentes de datos administrativos existentes es prometedora como una solución estadística para producir conteos de

¹⁹ Bycroft, C (2010). *A register-based census: what is the potential for New Zealand?* Wellington. Statistics New Zealand.

población que evitarían la necesidad de entrevistar a toda la población. Este es el enfoque actual de las investigaciones de Statistics New Zealand.

Durante las visitas a los países en entrevistas mantenidas con algunas entidades también se encontró que falta una visión integradora de las entidades (funciones) públicas. Las entidades no comparten información ni siquiera con fines de control administrativo, mucho menos para generar estadísticas integrando varias fuentes administrativas. Las entidades públicas no conocen la información que generan otras entidades y que podrían ser utilizadas con fines de control administrativo, como mínimo, al cruzar esta información entre sí.

Otros factores comunes a todos los países de la subregión son: factores culturales, pues no existe cultura de uso de RA con fines estadísticos, resistencia al cambio, desconfianza en los datos de los RA; pérdida de control de los procesos de diseño, recogida y procesamiento de datos; por otro lado existe gran interés en utilizar fuentes administrativas para elaborar estadísticas.

En cuanto a las capacidades institucionales, también se encuentran diferencias entre los países como lo marca la Tabla 3. Matriz integrada de evaluación de la capacidad institucional, y a partir del análisis de los informes de diagnóstico nacionales.

Las preguntas del cuestionario CECAP correspondientes a la dimensión Recursos Humanos, indagan sobre si el personal está capacitado en diferentes áreas/temas pero no evalúan si es suficiente, esto tal vez podría llevar a que este indicador esté sobrevaluado. Por ejemplo, en el caso de Bolivia el indicador de RRHH correspondiente a personas es 3 lo cual significa que el personal está capacitado en todas las áreas/temas indagados, pero como se señala en las propias observaciones del registro de personas de la matriz integrada de evaluación de la capacidad institucional del taller regional, se indica que el personal es escaso. Asimismo, en las observaciones de RRHH de inmuebles se señala que “cuenta con personal de planta capacitado en tecnologías de la información para transformar RA en RE”, pero en las observaciones de Recursos institucionales se indica que “Aunque cuenta con un sistema informático que le permite transformar la información del RA en estadístico, este presenta problemas por falta de personal”. Esto confirma que el indicador podría estar sobredimensionado, pues evalúa la capacitación del personal pero no si es el suficiente para ejecutar las actividades.

En los cuatro países los aspectos de la capacidad institucional que presentan mayores debilidades son Recursos institucionales y Control de calidad. Esto se debe a que no se cuenta con inventarios actualizados de RA y que los lineamientos y metodología establecidos por los INE no se han aplicado en las entidades del SEN. También podríamos agregar la dimensión de Recursos humanos, pues parecen no ser suficientes a pesar de que en la mayoría de los casos el personal está capacitado. Además, se presentan problemas debido a la alta rotación del personal y las formas contractuales.

En cuanto a la visión estratégica sobre el uso de RA con fines estadísticos también se evidencia una diferencia entre los países. En el caso del INE de Bolivia se denota más esta diferencia pues al momento de plantear su propuesta de prueba piloto del SIREPI existen muchas incertidumbres y solicitan a los otros INE, en particular al DANE como coordinador del proyecto, directivas sobre qué hacer. Surge la duda si el piloto del SIREPI es visualizado simplemente como algo que se debe cumplir en el marco del proyecto BPR, o si por el contrario, se entiende como un elemento clave para dar operatividad al plan estratégico del INE en cuanto a pruebas que la entidad debería

realizar en el marco de su visión estratégica sobre el uso RA para la generación de estadísticas como parte del plan de modernización en el que se ha embarcado la institución.

En cambio en el caso de Colombia y Ecuador la diferencia está determinada por las pruebas piloto que ya han realizado, pues en realidad avanzaron más no por cumplir con los objetivos del proyecto BPR en sí, sino por la necesidad de implementar la visión estratégica de la institución.

Sin embargo, el hecho que cada INE continúe avanzando por su cuenta, en algún caso en forma descoordinada del resto, no necesariamente alineado con los objetivos del proyecto o no siguiendo un plan de acción común, puede ser contraproducente para el proyecto en sí.

Es así que aparece el dilema si la metodología común (con adaptaciones a cada país) a desarrollarse en el marco de este proyecto se va a aplicar realmente por todos los INE o si alguno va a desarrollar la propia, no porque existan diferencias reales en el contexto de cada país, sino más bien porque quiere diferenciarse o destacarse del resto y tener su metodología y herramientas particulares con su sello propio.

Ya se han encontrado evidencias al respecto, pues los INE han creado o están en proceso de crear instrumentos prácticamente iguales a otros elaborados por organismos internacionales u otros países, pero con matices para darle el toque de distinción. No se comprende por qué se destinan recursos en “reinventar la rueda”.

La siguiente matriz explica a modo de resumen los factores comunes que impactan en la producción, acceso y uso de los registros administrativos en la subregión.

Tabla 4. Matriz de factores comunes que impactan en la producción, acceso y uso de los registros administrativos en la subregión.

Característica o problema	Causa (posible)
Acceso a los microdatos brutos identificados de los RA	Marco legal inadecuado
	Falta de acuerdos de intercambio de información con especificaciones técnicas
	Factores político-institucionales (*) (**)
Entorno institucional de los INE adecuado para la producción de estadísticas basadas en RA.	Falta de visión estratégica de desarrollo estadístico con énfasis en el uso de RA (a nivel país)
	Apoyo de la alta dirección que no se traduce en acciones concretas
	Estructura organizativa inadecuada
	Falta de recursos humanos capacitados y estables
Integración de RA	Falta de tecnologías y metodologías
	Falta de claves de identificación estandarizadas
	Problemas de interoperabilidad entre las entidades del SEN
	Las entidades no comparten información con fines de control administrativo
Organización y coordinación del trabajo	Falta de visión integradora de las entidades (funciones) públicas
	No hay un uso sistemático de inventarios de RA
	No se ha profundizado en la evaluación de la calidad de RA

	Falta retroalimentación a las entidades del SEN, acciones ganar-ganar (capacitación, mejora de la calidad de procesos y datos de RA)
	Coordinación del SEN inadecuada, contacto con entidades externas no centralizado ni coordinado.
Comunicación	Problemas de comunicación interna sobre uso de RA, proyecto BPR.
	Se trabaja en “silos” o “islas” y no se comparte información ni experiencias.

(*) Bolivia: los Gobiernos Autónomos Departamentales y Gobiernos Autónomos Municipales (establecidos en la constitución de 2010), argumentando su autonomía de gestión, podrían no alinearse con las estrategias nacionales en cuanto a la armonización y estandarización de ciertos registros administrativos clave como son los catastros municipales para integrarlos en un catastro único nacional.

(**) Ecuador: el contexto político-institucional actual ha facilitado el acceso del INE a los microdatos identificados de los RA de las fuentes administrativas. Surge la duda si se logra institucionalizar y garantizar su sostenibilidad en el futuro, incluso con un escenario político-institucional desfavorable.

Fuente: elaboración propia del consultor.

Factores técnicos comunes que podrían afectar la creación y sostenibilidad del SIREPI:

- Falta de clave estandarizada en RA de inmuebles.
- Domicilios no estandarizados.
- No existe registro de residencia consolidado y actualizado.
- No se evalúa la calidad de los datos en forma sistemática.
- Falta de estándares de calidad.
- Escasa interoperabilidad.
- Falta de experiencia y formación de técnicos en TI.
- Falta de infraestructura de TI.

Factores de éxito, fortalezas y oportunidades comunes a ser aprovechadas para la conformación del SIREPI.

- Contexto regional: Proyecto regional BPR.
- Apoyo de la alta dirección.
- Demandas de información no satisfechas (indicadores de los ODS).
- Proyectos de mejora de interoperabilidad de las entidades públicas.
- Experiencias de los países que ya han avanzado en pilotos en la subregión.
- Proyectos de integración de registros administrativos ya existentes (padrones integrados de beneficiarios de programas del gobierno).

El marco legal y entorno político-institucional de Ecuador han facilitado el acceso del INEC a los registros administrativos que ha solicitado a las entidades del SEN, sobre todo en el marco del ejercicio piloto del censo de Galápagos.

Esto sin dudas es un rasgo distintivo de este país con respecto al resto, quienes han tenido mayores dificultades para acceder a los microdatos identificados de los registros administrativos.

Seguramente esto allanará el camino del INEC hacia la creación de un sistema integrado de registros estadísticos, aunque el INEC está más enfocado en su objetivo principal que es realizar un Censo mixto en el 2020 (combinación de métodos tradicionales e información proveniente de registros administrativos) el cual sería el último censo realizado en forma tradicional, con una meta a más largo plazo para el 2030 donde esperan realizar un censo basado exclusivamente en registros administrativos.

Este contexto propicio debería ser aprovechado no sólo para lograr acceso a los datos administrativos de las entidades, sino también para institucionalizarlo y garantizar su sostenibilidad en el futuro, sobre todo si el escenario político-institucional dejase de ser tan favorable.

Cabe señalar que, si bien es fundamental contar con un marco legal adecuado, únicamente esto no garantiza el acceso a los registros administrativos y su uso con fines estadísticos. Por tanto, como se ha señalado en el párrafo anterior, se deberían tomar medidas adicionales para institucionalizarlo y garantizar su sostenibilidad.

Estas medidas van desde acuerdos interinstitucionales de cooperación e intercambio de información, capacitación, asistencia técnica, documentación de metadatos, elaboración de inventarios de registros administrativos hasta certificaciones de calidad y sello de estadística oficial por parte de los INE.

A pesar de las dificultades y los factores que impactan negativamente en la producción, acceso y uso de los registros administrativos en la subregión, se evidencia la voluntad de alcanzar los objetivos propuestos. Sin embargo, el grado de compromiso de cada INE, más bien de los funcionarios afectados al proyecto, es marcadamente diferente. También existen notorias diferencias en cuanto a la formación, experiencia y motivación del personal en relación a la producción de estadísticas basadas en registros administrativos.

Por tanto, se entiende necesario profundizar en la capacitación e intercambio de experiencias entre los países, sobre todo los que más han avanzado en el tema. Se podría pensar en capacitaciones en modalidad de taller o visitas de técnicos de un país a otro pero para trabajar en un tema concreto, para apoyar en la resolución de un asunto particular bien delimitado. En lugar de una capacitación genérica sobre uso de RA con fines estadísticos, pues ya se han realizado entrenamientos de este tipo en la región. Ahora se necesita algo más concreto y con impacto en el corto plazo.

Para el diseño de este plan de capacitación y apoyo entre países es fundamental tomar el marco conceptual y metodológico que se elaborará en el marco de esta consultoría como guía para determinar los puntos fuertes y las necesidades de capacitación o de apoyo a cada país. De esta manera se podrá establecer qué INE estaría en condiciones de apoyar a otros en determinados temas y cuáles recibirían la asistencia técnica. Asimismo se podrían identificar las áreas que necesitan apoyo en los cuatro INE, por lo

cual requerirán otro tipo de asistencia técnica (consultores, países fuera de la subregión).

Interoperabilidad

Una de las definiciones más amplia y completa en la actualidad establece que la interoperabilidad es “la habilidad de organizaciones y sistemas dispares y diversos para interactuar con objetivos consensuados y comunes y con la finalidad de obtener beneficios mutuos. La interacción implica que las organizaciones involucradas compartan información y conocimiento a través de sus procesos de negocio, mediante el intercambio de datos entre sus respectivos sistemas de tecnología de la información y las comunicaciones”²⁰.

Este tema ha debido ser analizado en profundidad durante las visitas a los países y con la información solicitada en las fichas enviadas a cada INE, puesto que no se han encontrado en los informes de diagnósticos nacionales elementos suficientes para el análisis sobre la temática. Esto también se señala como una propuesta de mejora en el informe de diagnóstico de Perú. En el informe de diagnóstico de los RA de población de Ecuador se han incluido algunos elementos de análisis sobre la interoperabilidad de los RA, siendo el informe que contiene el mayor grado de profundidad en la evaluación de este aspecto.

Si bien la matriz integrada de “Condiciones para la conformación del SIREPI y para transformar registros administrativos en registros estadísticos” contempla en parte estos aspectos, es necesario llegar al detalle en ciertos puntos:

¿Qué RA se pueden vincular y a través de qué variables (clave o no)?

¿Cuáles son los archivos de microdatos de los RA que el INE tiene acceso?

¿El INE tiene acceso en forma periódica a estos archivos?

¿Es necesario que el INE reitere la solicitud de autorización de acceso a los archivos en cada oportunidad?

¿El intercambio de datos/archivos se hace a través de qué mecanismo/formato o es un sistema integrado? ¿Cómo funciona? ¿Está documentado?

En el caso de Perú, se ha podido tener acceso a un informe reciente de diagnóstico de la situación de la interoperabilidad en entidades del estado peruano²¹, donde se destacan las siguientes conclusiones generales del estudio:

- “No hay consenso en el uso de una metodología única de proyectos de interoperabilidad, no hay una guía para la implementación de un Marco de Interoperabilidad único.

²⁰ Bases para una estrategia iberoamericana de interoperabilidad. Documento marco iberoamericano de interoperabilidad ratificado en 2010 por la XX Cumbre Iberoamericana de Jefes de Estado y de Gobierno.

²¹ Consejo Nacional de la Competitividad (2015). *Diagnóstico de la situación de la interoperabilidad en entidades del estado peruano*. Informe final elaborado por el consultor Alberto Ricardo Cáceda Farfán, enero de 2015, Lima – Perú.

<http://www.cnc.gob.pe/images/upload/paginaweb/archivo/25/Diagn%C3%B3stico%20de%20la%20Interoperabilidad%20en%20Entidades%20del%20Estado%20Peruano.pdf>. Último acceso 20-11-2016.

- Falta de aplicación de la normatividad en estandarización de protocolos y lenguaje para el intercambio de información, siendo notoria la ausencia de una institución que convoque y supervise el desarrollo de la semántica en la interoperabilidad a nivel Estado.
- Falta de un marco legal que posibilite el uso y consumo de servicios de intercambio de información.
- Falta el rol proactivo de una entidad responsable en el tema, a efectos de mejorar los niveles de articulación y coordinación a nivel intra-institucional (dentro de cada entidad) e interinstitucional.
- Falta de política de intercambio de datos que haga más sencillo el proceso de poder acceder a la información que manejan las diferentes entidades del estado, desarrollando la cooperación interinstitucional y evitando los lentos convenios de intercambio de datos con los cuales se procede actualmente.
- La falta de institucionalización a nivel país del tema de gobierno electrónico e interoperabilidad afecta gravemente su avance, razón por la cual se requiere hacer ajustes en términos de legislación, planificación, gestión, monitoreo, evaluación, organización, entre otros.
- No existe en el país una política clara que norme la interoperabilidad y que conduzca el desarrollo del proceso, tal como existe en otros países.”

En este mismo informe se incluye un análisis acerca de la situación de Colombia sobre interoperabilidad dentro de un capítulo de experiencias internacionales relevantes en este tema. Donde se señala que es uno de los países que más ha avanzado hacia el desarrollo de una estrategia de Gobierno Electrónico, ubicándose en los primeros lugares en rankings de los indicadores de Gobierno Electrónico y Servicios en línea de Naciones Unidas. Lo cual se traducirá en beneficios importantes para el DANE en el marco de este proyecto.

En tanto Ecuador ha definido una serie de proyectos orientados a consolidar la interoperabilidad en las entidades del estado, estableciendo las siguientes acciones estratégicas dentro del Plan Nacional de Gobierno Electrónico 2014-2017:

- “Puesta en marcha y proyección de la interoperabilidad en Ecuador.
- Consolidación y robustecimiento del Bus de Servicios Gubernamentales.
- Norma y portal de interoperabilidad (tecnológica, organizacional y semántica) y acceso y uso al BSG.
- Ejecución del Comité de Interoperabilidad (coordinación de Bus de datos, Bus Transaccional y Bus Financiero).
- Comunicaciones unificadas” ²².

Sistema de registros

La evaluación de la calidad de los RA se realizó en forma individual, registro a registro, o al menos por entidad responsable de uno o varios RA, pero no se efectuó un análisis global considerando los RA evaluados como un todo.

En este sentido, durante las visitas a cada país fue necesario evaluar los RA como parte de un sistema integrado (visión a futuro), determinando claramente las variables clave de identificación comunes, variables a utilizar de cada RA con mejor calidad, variables

²² Interoperabilidad y Seguridad. Secretaría Nacional de la Administración Pública. Ecuador.
<http://www.administracionpublica.gob.ec/interoperabilidad-y-seguridad>. Último acceso 20-11-2016.

de validación o control y variables que indiquen la fecha de actualización de los datos provenientes de cada RA.

La evaluación de la calidad de los registros administrativos deberá realizarse considerando la integración o vinculación de varios RA en lugar de evaluar la calidad de un solo RA a la vez. Al integrar varios RA se logra mejorar la calidad de los datos, cobertura y ampliar el análisis de la información estadística producida a partir de éstos.

Es necesario, entonces, incluir en la evaluación no sólo los RA a los que se tiene acceso actualmente sino a todos aquellos RA que sean necesarios para el fortalecimiento del sistema y mejora de la calidad de los datos y cobertura del registro estadístico.

Qué registros administrativos incluir en el sistema dependerá de las necesidades de información y de la calidad y cobertura de los registros administrativos en su globalidad como se ha mencionado anteriormente.

Por tanto, es primordial identificar las necesidades de información (partir desde la demanda) y contrastarlas con la información disponible en los inventarios de registros administrativos (oferta). De esta forma se podrá priorizar el acceso a los registros administrativos necesarios.

Se recomienda establecer etapas para la implementación del sistema, luego definir objetivos por cada etapa y para cada uno de estos objetivos definir la información que se pretende obtener a partir del sistema.

Esto ayudará, además, a vincular indicadores con fuentes administrativas hasta nivel de variable (metadatos, trazabilidad), priorizar los registros administrativos a los que se debe acceder y trazar un plan de acción.

Marco legal y mecanismos de coordinación

De acuerdo a la “Matriz integrada sobre mecanismos actuales de coordinación” los cuatro países declaran que los INE tienen mecanismos de coordinación con las entidades responsables de los RA.

En muy pocos casos se especifica que el “convenio interinstitucional” implica la transferencia de bases de datos al INE.

La normativa vigente o las nuevas leyes que se están planificando en los países carecen de detalles técnicos como es de esperarse, por lo tanto deben acompañarse de reglamentaciones para su adecuada aplicación.

Estas reglamentaciones o decretos complementarios deberían incluir ciertos aspectos técnicos tales como la interoperabilidad o transferencia de bases de datos al INE, para garantizar la continuidad del acceso a la información y minimizar la cantidad de convenios necesarios con cada una de las entidades responsables de los registros administrativos, pues estos estarían contemplados, o al menos una gran parte, en la reglamentación o decretos complementarios de las leyes.

Entretanto no se disponga del marco legal y reglamentaciones adecuados los INE seguramente recurrirán a tantos convenios como registros administrativos sean requeridos. Entonces, se debería pensar en un formato de convenio estandarizado que se pudiera aplicar en todos los casos con ajustes mínimos para contemplar las particularidades de cada asunto.

Es indudable que el marco legal, incluida su reglamentación, deberá ajustarse a las necesidades actuales y futuras que exige un sistema integrado de registros estadísticos, en el sentido de interoperabilidad, estandarización de claves, interconexión de bases de datos, uso de clasificadores estándar, validación de datos entre diferentes sistemas de diferentes entidades, entre otros.

El marco legal por sí solo no es suficiente, pues debe ser acompañado de mecanismos de coordinación interinstitucional liderados por el INE que aseguren la mejora continua del sistema integrado de registros estadísticos y para revisar la reglamentación o decretos complementarios para ajustarlos a las necesidades cambiantes del sistema.

Los consejos o comisiones de estadística integradas por los miembros del SEN (en algunos casos establecidos por ley), no siempre han funcionado según lo esperado, pero en el marco del sistema integrado de registros estadísticos cobran un rol mucho más preponderante como mecanismo de coordinación y facilitador del entorno de interoperabilidad necesario.

Lecciones aprendidas y buenas prácticas

- El DANE de Colombia ha estado aplicando herramientas de modelado de procesos utilizando el estándar BPMN para diseñar y documentar los procesos de integración de registros y transformación de RA en registros estadísticos.
- El desarrollo de algoritmos para la unión probabilística de registros ha sido uno de los avances importantes que ha realizado el DANE de Colombia. Las herramientas de software más comúnmente utilizadas para efectuar la unión de registros por métodos probabilísticos proveen de algoritmos estándar para ejecutar este procedimiento, pero no siempre se ajustan a las necesidades de los usuarios. Sería recomendable profundizar en pruebas y comparar resultados entre los algoritmos estándar de la industria del software y los algoritmos desarrollados por el DANE, incluso aplicándolos en los cuatro países de la región.
- El INEC de Ecuador está utilizando herramientas de Big Data para el procesamiento y análisis de grandes volúmenes información de RA.
- Es necesario documentar el proceso de transformación de los RA en registros estadísticos, mapear las variables del RE con las variables originales de los RA y asociar las listas de categorías de las variables del RE con las de los RA. En este sentido, el INEC de Ecuador ha desarrollado la herramienta de documentación de metadatos Metadec, que cubre parcialmente esta necesidad, pues permite asociar los catálogos o listas de categorías de variables del RE con las correspondientes en los RA.
- Se están dando los primeros pasos hacia el desarrollo de herramientas de evaluación de la calidad de un sistema integrado de registros estadísticos. El DANE de Colombia e INEC de Ecuador han venido trabajando en una serie de instrumentos para evaluar la calidad de los RA. Es necesario profundizar en el desarrollo de una herramienta de evaluación integral, no sólo orientada a evaluar RA a RA en forma individual, sino más bien orientada al sistema integrado de registros estadísticos. Pues como se ha mencionado anteriormente, lo que se busca con la integración de varios RA, además de mejorar la cobertura, es mejorar la calidad de los datos, utilizando aquellas variables consideradas de mayor calidad.

- Los INE poco a poco se están adecuando a los nuevos procesos de producción estadística basada en RA. En el caso del INEC de Ecuador se ha creado recientemente una dirección especializada en el uso de RA con fines estadísticos. El INE de Bolivia está inmerso en un proyecto de modernización liderado por la Unidad de modernización para mejorar la gestión del INE con base en el uso de RA.
- Avanzar con lo que se tiene. Generar estadísticas basadas en RA a partir de la información que se dispone actualmente sobre un sector o temática en particular, pero con una visión integradora a largo plazo. Un buen ejemplo de esto es lo que ha realizado el INEI de Perú acerca de estadísticas sobre violencia y hechos delictivos.
- Es necesario contar con un catastro a nivel nacional al menos en lo que se refiere al establecimiento de normas técnicas, protocolos y números de identificación únicos a nivel nacional con claves estandarizadas. También para evitar omisiones o solapamientos de predios especialmente en los límites municipales.
- Integración de varias áreas, creación de grupos de trabajo interdisciplinarios, mejorar la comunicación interna. Es clave socializar los avances del proyecto y el plan de acción, fomentar el trabajo en equipo y la integración de varias áreas, sobre todo las áreas temáticas o productoras. Esto también ayudará a vencer la natural resistencia al cambio que se genera al interior de las organizaciones cuando se promueven proyectos de alto impacto tanto en los procesos como en los productos que se generan.
- Actualización del marco legal para adecuarlo a las necesidades y contexto actuales. Todos los países están en este proceso.

Recomendaciones:

Las siguientes recomendaciones no se presentan en ningún orden en particular, simplemente se ha enumerado una serie de acciones que podrían coadyuvar a la implementación del SIREPI, con base en las experiencias de los países de la CAN, buenas prácticas internacionales y experiencia del consultor.

- **Crear o mejorar el Catastro Nacional**

Excepto Colombia, el resto de los países cuentan con catastros descentralizados a nivel municipal. En algunos casos ya se han iniciado proyectos para desarrollar un catastro nacional. Lo importante es mejorar la clave de identificación de predios, logrando su estandarización y evitando duplicidades.

Este catastro nacional estaría encargado de establecer normas técnicas, protocolos y estandarización de las claves de identificación de predios.

Además, trabajaría en minimizar las omisiones o solapamientos de predios que se producen entre municipios colindantes.

- **Uso de registros administrativos de servicios públicos**

Se ha mencionado anteriormente que los registros administrativos de inmuebles presentan deficiencias en cuanto a las claves de identificación, por lo cual se debe recurrir a la variable de domicilio como clave de identificación y vinculación entre RA de inmuebles y entre éstos y los RA de población, pues en este último caso no se

dispone de la clave código o número de padrón del inmueble donde reside la persona.

Se ha constatado que los RA de servicios públicos, en especial el servicio de energía eléctrica, presenta una buena cobertura y actualización de los datos, contando con la variable de domicilio e incluso las coordenadas geográficas de la ubicación del medidor del consumo.

Es por esta razón que se recomienda su uso siempre que sea posible, para vincular registros de inmuebles entre sí y con RA de población.

- **Documentar metadatos de los RA y RE**

Los metadatos de las fuentes de información son un elemento clave para lograr un mejor aprovechamiento estadístico de las mismas.

En este sentido se recomienda utilizar el estándar internacional DDI para documentar los metadatos de los RA. De esta forma se logrará uniformizar la forma de documentar y se evitarán vacíos de información en áreas importantes o documentación excesiva en temas irrelevantes.

Asimismo, es fundamental documentar el proceso de transformación de los RA en registros estadísticos, mapear las variables del RE vinculándolas con las variables originales y sus fuentes administrativas y registrar la asociación entre los catálogos o categorías de las variables del RE (estandarizadas de acuerdo a las normas técnicas del INE) y las variables originales de los RA.

- **Crear grupos de trabajo interdisciplinarios**

Se han problemas de comunicación interna en los INE que podrían afectar la implementación del SIREPI. Por lo cual, se recomienda crear grupos de trabajo conformados por técnicos de varias áreas para abordar el tema del uso de registros administrativos con fines estadísticos, implementación del SIREPI y cómo va a impactar esto en los procesos y estructura organizativa del INE.

También se debería trabajar en comunicar mejor los siguientes pasos, plan de acción y etapas, objetivos y resultados esperados del proyecto.

- **Facilitar la cooperación sur-sur**

Realización de talleres dictados por los INE que mayor experticia tienen en determinados temas y pasantías e intercambio de técnicos sobre temas clave.

- **Crear y mantener un inventario de registros administrativos**

Uno de los primeros pasos hacia la construcción de un sistema integrado de registros estadísticos es elaborar un inventario de registros administrativos para conocer las fuentes de información administrativa disponibles en el SEN que podrían ser de utilidad para elaborar estadísticas oficiales.

La mayoría de los países ha conducido procedimientos para obtener información sobre las operaciones estadísticas que realizan las entidades del SEN, ya sea en el marco del diseño de los planes o estrategias nacionales de desarrollo estadístico (con un propósito puntual), o con una visión a largo plazo de mantener un inventario actualizado de operaciones estadísticas.

En cualquiera de los dos casos, el listado o inventario de operaciones estadísticas tiene por objetivo registrar la información estadística que generan actualmente las entidades del SEN por medio de operaciones estadísticas, que en la mayoría de los casos están basadas en registros administrativos. Obviamente este trabajo debería ser la base o el punto de partida para la conformación del inventario de registros administrativos.

Es recomendable que los cuatro países utilicen herramientas estandarizadas y armonizadas. Es en este tipo de elementos que es posible pensar en armonización y estandarización de procesos y metodologías entre los países de la subregión.

- **Evaluaciones de la calidad exhaustivas. Evaluar la calidad de los datos del RA**

En este informe, especialmente en el anexo I.1, se ha señalado algunas observaciones acerca de los instrumentos que se han utilizado para evaluar la calidad de los RA. Se ha destacado la falta de exhaustividad en la evaluación de la calidad de los datos de los RA, en particular la dimensión de precisión y exactitud que evalúa muy indirectamente (o casi sin relación alguna) la calidad del dato en sí mismo.

Si bien la calidad del dato del RA está determinada por la calidad de los procesos, y es correcto incluir en las evaluaciones dimensiones de la calidad del proceso, en algún momento es necesario evaluar más precisamente la calidad del dato del RA. Más aun considerando que los datos no fueron capturados por el INE y los instrumentos utilizados no fueron diseñados con fines estadísticos.

Por tanto, las evaluaciones de la calidad del RA deberían incluir indicadores directos de la calidad de los datos del registro.

Por mencionar algunos de ellos:

- Tasa de cobertura.
- Tasa de respuesta (global y por pregunta).
- Tasa de errores de clasificación (para variables que usen clasificadores).
- Lapso entre el período de referencia de los datos y la fecha de publicación de los principales resultados estadísticos (oportunidad).
- Porcentaje de unidades o casos que pertenecen a la población objetivo creadas durante el año "t" que han sido registradas antes del fin del año t+1 (actualidad de los datos).
- Tasa de registros (unidades, casos, filas) con clave única.
- Tasa de registros múltiples de la misma unidad.

- **Evaluaciones globales de calidad del sistema integrado de RE**

Se recomienda realizar evaluaciones integrales de la calidad, no sólo orientadas a evaluar RA a RA en forma individual, sino más bien orientadas al sistema integrado de registros estadísticos.

Lo que se busca con la integración de varios RA, además de mejorar la cobertura, es mejorar la calidad de los datos, utilizando aquellas variables consideradas de mayor calidad.

Es muy importante evaluar la calidad de los RA individualmente, pero sin dudas al integrar varios RA en uno o varios RE la calidad global va a cambiar con respecto a

las evaluaciones individuales, pues esto es uno de los resultados que se busca con la integración de varios RA.

Tal vez un RA es considerado de baja calidad en una evaluación individual, pero al integrarlo con otros RA pues se considera que tiene buena cobertura, se logra una calidad global del RE aceptable.

- **Implementar acciones ganar-ganar con las fuentes administrativas**

Los INE deberían diseñar estrategias basadas en acciones ganar-ganar con las entidades del SEN con el fin de asegurar el acceso a los datos administrativos por medio de relaciones sólidas sustentables a largo plazo.

Como forma de retribución “a cambio” de los datos administrativos, los INE podrían ofrecer a las fuentes administrativas capacitación, asistencia técnica para la mejora de la calidad de los procesos y los datos de los RA y/o generar estadísticas de acuerdo a las necesidades de cada entidad.

- **Desarrollar estrategias a corto y largo plazo, implementación por etapas**

Es recomendable definir un plan de acción por etapas con objetivos bien definidos. Asociar estos objetivos con indicadores o estadísticas a generar en cada etapa. Mapear los indicadores con variables y sus fuentes administrativas.

Se deberían establecer etapas para la implementación del sistema, luego definir objetivos por cada etapa y para cada uno de estos objetivos definir la información que se pretende obtener a partir del sistema.

Esto ayudará, además, a vincular (mapear) indicadores estadísticos con fuentes administrativas hasta nivel de variable (metadatos, trazabilidad), priorizar los registros administrativos a los que se debe acceder y trazar un plan de acción.

Este plan de acción debería estar basado en estrategias a corto y largo plazo de acuerdo a las posibilidades y contexto de cada país.

Algunas estrategias a corto plazo:

- Crear/actualizar el inventario de registros administrativos
- Documentar metadatos de los RA
- Establecer convenios con las fuentes administrativas
- Realizar pruebas piloto de integración de RA
- Generar estadísticas basadas en RA para sectores específicos a partir de los RA a los que se tiene acceso actualmente.
- Elaborar plan de evaluación de la calidad de los RA

Estrategias a largo plazo:

- Actualizar el marco legal
- Mejorar la interoperabilidad entre las entidades del SEN
- Acuerdos marco entre todas las entidades del SEN para compartir RA
- Desarrollar un registro de residencia
- Crear/mejorar el catastro nacional
- Implementar el SIREPI

- **Implementación del SIREPI adaptada al contexto de cada INE**

Es evidente que la implementación del SIREPI deberá adaptarse a las posibilidades de cada INE y al contexto del país.

En este proceso, algunos países intentarán llegar a la integración de RA y probablemente lo logren, mientras otros lograrán cimentar las bases para la construcción del sistema integrado de registros estadísticos.

Se hace hincapié en la recomendación anterior acerca de la implementación por etapas basada en estrategias a corto y largo plazo, adaptándola al contexto de cada país, desarrollando un plan de acción que pueda ser cumplido por cada INE.

La estrategia utilizada por Nueva Zelanda podría ser también de utilidad para la región, es decir, usar los RA disponibles actualmente y a largo plazo pensar en un registro de residencia como han desarrollado los países escandinavos. Pero siempre teniendo en cuenta la realidad de cada país.

- **Diseño del SIREPI**

Se recomienda diseñar el SIREPI pensando en un modelo de gestión basado en tres pilares fundamentales: personas, procesos y tecnología, es decir, que el diseño del SIREPI no esté enfocado en la tecnología exclusivamente, sino que además esté alineado con los procesos y éstos con la estrategia.

En cuanto a los aspectos tecnológicos, se debería considerar un sistema flexible, parametrizable, lo que comúnmente se llama software generalizado. Deberá ser escalable, para que acompañe el crecimiento del proyecto en cuanto a cantidad de RA, volumen de información, capacidad de procesamiento; y multiplataforma (para facilitar la interoperabilidad), lo que permitirá funcionar en diversas arquitecturas tecnológicas, sistemas operativos y hardware.

En resumen, se recomienda utilizar herramientas de software denominadas ETL (por su sigla en inglés Extract, Transform and Load, o Extracción, Transformación y Carga) para la recopilación de datos de diferentes fuentes y formatos (extracción), la depuración, estandarización de variables e integración de registros (transformación) y la carga de datos en la base de datos, datawarehouse o datamart (registro estadístico).

Existen varias herramientas ETL en el mercado, dentro de las open source encontramos a una de las más populares como Pentaho Data Integration, y además la suite Pentaho dispone de herramientas de Business Intelligence y Business Analytics para el análisis y explotación de datos.

Utilizar herramientas ETL open source permitiría iniciar en forma gradual, por etapas sin necesidad de realizar una gran inversión al inicio. Luego, cada INE podría determinar si continúa con la versión “Community” open source o si en cambio invierte en la versión empresarial de pago.

- **Aprovechar proyectos de integración de RA ya existentes en el país**

Es importante ahondar en proyectos de integración de registros administrativos ya existentes o que cuentan con apoyo de otras entidades como son los padrones de beneficiarios de programas del gobierno, pues su sustentabilidad, interoperabilidad y acceso a los datos está asegurada por un marco legal e instituciones con mayor peso y poder político-institucional.

Experiencias en la región:

En la región podemos destacar dos casos exitosos de sistemas integrados de registros estadísticos de población basados en padrones de beneficiarios de programas del gobierno, como son el RIS (Registro de Información Social) implementado por el Ministerio de Desarrollo Social de Chile y el SIIAS²³ (Sistema Integrado de Información del Área Social) del Ministerio de Desarrollo Social de Uruguay. Este último integra más de 80 registros administrativos de población e inmuebles provenientes de diversas entidades²⁴. Inicialmente, estos sistemas fueron concebidos para cumplir con el requerimiento administrativo de contar con información integral, global sobre las prestaciones sociales que recibe cada persona y así los diferentes organismos pueden consultar el sistema y tomar decisiones informadas. Luego, a partir de este gran registro administrativo se logró conformar un sistema integrado de registros estadísticos de personas, lo que ha permitido generar una muy importante serie de indicadores sociales a muy bajo costo, con una gran cobertura y actualización periódica.

En ambos casos los Ministerios responsables han logrado la colaboración de las entidades que comparten la información de sus registros administrativos. La clave de este logro se basa en que el sistema integrado de información es de utilidad para todas las entidades que forman parte del proyecto, pues les brinda información sobre las prestaciones que recibe cada beneficiario a través de los diferentes programas de gobierno. Es así que cada entidad se preocupa de mantenerlo actualizado. Obviamente esto facilita la generación de un sistema integrado de registros estadísticos de población, además que ya se han sorteado los escollos inherentes a compatibilidad de sistemas informáticos, bases de datos, claves de identificación única de registros, etc. Es así que los INE pueden beneficiarse de estos proyectos para allanar el camino hacia un sistema integrado de registros estadísticos.

Otros casos de éxito los podemos encontrar en los estados de Yucatán y Guanajuato, México, donde las entidades de planeación han liderado proyectos para la creación de sistemas integrados de registros estadísticos. En ambos casos el enfoque fue algo diferente, pues las entidades de planeación y presupuesto tienen más poder o peso para obtener información de las fuentes administrativas y lograr su colaboración. En estos ejemplos se definieron una serie de indicadores de las MIR y para cada indicador se hizo el ejercicio de identificar las posibles fuentes de datos administrativos (se partió desde la demanda de información) lo cual facilita la búsqueda de las fuentes de información y se cubre la demanda.

En los cuatro ejemplos citados, la implementación del sistema integrado de registros estadísticos se apoyó estratégicamente en los proyectos de integración de registros administrativos liderados por las entidades mencionadas (Ministerio de Desarrollo Social y entidades de planeación), pues eso facilitó la obtención de la información y la colaboración permanente y sistemática de las entidades que son fuentes de información administrativa.

Puesto que en general no existe aún la cultura de estadísticas basadas en registros y no siempre se logra la colaboración de las entidades, es importante buscar diferentes estrategias que faciliten obtener la información a partir de los registros administrativos.

²³ <http://siias.mides.gub.uy/>

²⁴ <http://siias.mides.gub.uy/innovaportal/v/59845/12/innova.front/listado-de-organismos-y-programas-en-el-siias>

Siempre que sea posible, es recomendable apoyarse en proyectos similares, si es que se han implementado o se van a implementar en el corto plazo en cada país. De lo contrario, dependerá del entorno institucional, marco legal, jerarquía o poder del INE de cada país.

En el caso de Colombia existe el proyecto SISBEN (Sistema de Selección de Beneficiarios Para Programas Sociales), en Ecuador el RIPS (Registro Interconectado de Programas Sociales) y en Perú el SISFOH (Sistema de Focalización de Hogares). Son sistemas integrados de beneficiarios de programas sociales que vinculan o están en proceso de integración de varias fuentes administradas sobre población y hogares. Es una buena oportunidad para que los INE aprovechen estas iniciativas y participen activamente de estos proyectos para que los RA también sean utilizados con fines estadísticos y formen parte del sistema integrado de registros estadísticos.

Anexo I.1 – Observaciones sobre los instrumentos de evaluación utilizados

A continuación se detallan ciertas consideraciones y propuestas de mejora de los instrumentos de evaluación de los RA.

Como observación general acerca de los instrumentos de evaluación CECAP y CECRA cabe mencionar que el cálculo de los indicadores agregados correspondientes a cada criterio o dimensión de la calidad está basado en un promedio simple de los indicadores (respuestas de las preguntas de los cuestionarios) que componen cada una de estas dimensiones. Esto implica que cada indicador dentro de cada dimensión tiene la misma ponderación, cuando en los hechos esto no es así, pues los modelos de evaluación multicriterio se caracterizan por estar conformados por atributos con diversos grados de importancia, según el grupo de interés, hay indicadores (preguntas) que tienen un mayor peso que otros en su dimensión.

La pregunta que surge entonces es ¿qué peso se le asigna a cada indicador? Existen varios métodos de ponderación de atributos cualitativos que se podrían utilizar para determinar dichos pesos (Diakoulaki, Entropía, Ordenación Simple, SMART y el Proceso Analítico Jerárquico, basados en preferencias o valoraciones de los usuarios sobre cada atributo; y las técnicas de análisis multivariado basadas en los datos ya recabados: Escalamiento Multidimensional o Análisis de Proximidades, Análisis de Componentes Principales, Análisis Factorial, Análisis de Correspondencias, Análisis de Correspondencias Múltiples.). Por citar un ejemplo, la Herramienta de Evaluación de la Calidad de los Registros Administrativos ²⁵ (HECRA) del Banco Mundial utiliza la técnica de análisis multivariado denominada Análisis de Componentes Principales ²⁶ para la ponderación de los atributos que componen cada dimensión o criterio de calidad.

También es oportuno señalar que los instrumentos **no realizan una evaluación sobre la calidad de los datos** de los archivos o base de datos del RA. La dimensión de precisión y exactitud se enfoca más bien en ciertos procesos o en la aplicación de tecnologías de la información que no necesariamente implica que se logre una mayor precisión y exactitud de los datos. Este elemento clave deberá ser incluido en alguna etapa del proceso de conformación del sistema integrado de registros estadísticos, pues resulta imprescindible evaluar la calidad de los datos de los registros administrativos que formarán parte de los registros base y otros registros estadísticos del sistema. Obviamente, esta evaluación se debería realizar en forma temprana, en etapas previas a la transformación e integración de RA en registros estadísticos, mejor aún si se hace inmediatamente después o junto con la caracterización de los RA (levantamiento del inventario de RA).

Seguramente no se contó con tiempo suficiente como para probar debidamente los instrumentos de evaluación antes de su aplicación en los diagnósticos nacionales, por medio de un piloto y taller de evaluación y ajuste de las herramientas para adecuarlas a los fines esperados.

²⁵ Banco Mundial. *“Guía de Auto-Diagnóstico de la Calidad de Registros Administrativos a ser utilizados con fines estadísticos”*. México, 2012

²⁶ Peña, D. *Análisis de datos multivariantes*. McGraw Hill, 2002.

Accesibilidad

“La Accesibilidad contempla la idoneidad en que están disponibles los datos, metadatos, los medios de divulgación, las metodologías y servicios de apoyo al usuario.” Parece ser que se refiere a datos estadísticos y usuarios de información estadística.

Preguntas que forman parte del criterio de “Accesibilidad”

36. Existe algún Convenio de entrega de información del RA a los usuarios y se cumple

36.A. A través de qué medio se entrega la información del Registro Administrativo

36.B. A través de qué formato se entrega la información del Registro Administrativo

36.C. Indique si se cuenta con alguna Normativa que permita a los usuarios el acceso a la información del Registro Administrativo

36.D. Indique si está disponible la información del Registro Administrativo hasta nivel de microdato o a qué nivel se encuentra disponible la información del RA

39. Se informa a los usuarios de TODOS los cambios y/o actualizaciones a los RA

40. Se comunica siempre a los usuarios los cambios y/o actualizaciones en las metodologías

No está claro si es accesibilidad a las estadísticas o al archivo bruto del RA, si es accesibilidad por parte de los usuarios primarios (INE) o los usuarios finales de las estadísticas basadas en los RA.

Un RA puede tener la mejor calificación en cuanto a accesibilidad, pues los informantes respondieron que se tiene total acceso a los datos, pero se debe a que pueden acceder fácilmente a las estadísticas, no así a los microdatos o base de datos bruta del RA.

Sobre la pregunta 36. “Existe algún Convenio de entrega de información del RA a los usuarios y se cumple. Se requiere saber con esta pregunta si existen por escrito términos de entrega de los datos entre la Fuente Administrativa y el usuario de la información del RA. Los términos de entrega deben incluir, como mínimo, una referencia al protocolo de intercambio de información, áreas responsables de la entrega y recepción de los datos, área responsable de la preparación de la información solicitada, calendario de entregas. Por ejemplo, los términos de entrega pueden ser parte de un acuerdo marco o un documento particular firmado entre la Fuente Administrativa y el usuario de la información del RA”.

Cuando habla del usuario del RA no queda claro si es un usuario como el INE o una unidad/depto. Interno de la misma entidad que lo va a explotar con fines estadísticos, entonces no queda claro si son datos de RA brutos, originales o si en cambio son datos estadísticos ya procesados.

Preguntas 36A a 36D es el mismo problema, no queda claro de qué usuario se habla, ni de qué datos (brutos, originales o datos estadísticos).

En la guía de llenado del cuestionario menciona que las preguntas 39 y 40 evalúan el criterio de calidad sobre la “credibilidad”, sin embargo se ha considerado en el puntaje del criterio de “accesibilidad”.

40. “Se comunica siempre a los usuarios los cambios y/o actualizaciones en las metodologías.”

“Esta pregunta está relacionada con la pregunta anterior y ambas evalúan el criterio de calidad sobre la credibilidad del Registro Administrativo.”

Precisión y exactitud

No hay ninguna pregunta/indicador referida a este criterio en forma directa, las evaluaciones de este criterio están basadas en indicadores indirectos o muy indirectos. Más asociados con procesos, pero que ni siquiera son procesos que podrían afectar directamente este criterio. Están más asociados a infraestructura de tecnologías de la información y bases de datos, que si bien tienen cierto impacto en la precisión y exactitud de los datos, pero en realidad hay otros procesos y factores clave que tienen un mayor impacto (o al menos más directo) en la calidad de los datos, o son fuentes de error mucho más determinantes que el propio sistema informático o la base de datos que soporta el RA.

Factores como la recogida de datos, si se hace con un entrevistador o si es auto-llenado, si el entrevistador recibió capacitación, si hay una revisión de la capacitación, si los formatos son ingresados al sistema en el mismo momento que se recibe la información o se produce el evento, si hay control de la digitación y verificación de la codificación manual, etc., son bastante más relevantes como fuentes de error y afectan más directamente la calidad del dato. Sin embargo no han sido considerados como indicadores de esta dimensión.

Podemos encontrar varios ejemplos donde el RA utiliza un sistema informático como soporte de su gestión y una base de datos para almacenar su información, pero los controles de calidad o reglas de validación de los datos no han sido validadas, o no están documentadas, o los cambios no son registrados/documentados, y esto sí afecta directamente la precisión de los datos, sin embargo no han sido tomados en cuenta para este criterio o dimensión de la calidad.

Se puede tener el mejor sistema informático pero si se le ingresan datos de mala calidad seguramente van a salir datos de mala calidad.

Se deberían incluir criterios como cobertura, tasa de respuesta, tasa de imputación y edición por variable, tasa de errores de clasificación (para variables que usen clasificadores), porcentaje de unidades o casos que pertenecen a la población objetivo creadas durante el año “t” que han sido registradas antes del fin del año t+1 (actualidad de los datos), tasa de registros (unidades, casos, filas) con clave única, tasa de registros múltiples de la misma unidad.

Es cierto que para obtener estos indicadores es necesario acceder a la base de datos del RA o llevar un registro de esta información a lo largo del tiempo. Esto no siempre se tiene o no se actualiza periódicamente. En su lugar, se podrían incluir indicadores indirectos de la precisión y exactitud asociados a los procesos como validaciones, controles de calidad y consistencia de la información, procedimientos, verificación de la digitación y codificación/clasificación, etc. Esto se señala también como una propuesta

de mejora en los informes de diagnóstico de los RA de población de Ecuador y de inmuebles de Colombia.

La única pregunta sobre cobertura que se incluyó fue la número 5, sobre cobertura geográfica, que en realidad no se puede generar un indicador a partir de ésta. Sin embargo, en algunos RA considerados en las evaluaciones nacionales sí aparece el dato de cobertura (en cuanto al universo de casos del registro), lo cual hace pensar que este dato podría ser calculado o estimado sin mayor dificultad para ciertos RA. Este sí es un atributo relevante del criterio de precisión y exactitud que se podría haber incluido en la evaluación.

La tasa de no respuesta se incluyó en el criterio de credibilidad, cuando en realidad corresponde al criterio de precisión y exactitud. Podemos encontrar muchos ejemplos de RA con una tasa de no respuesta inferior al 3%, sin embargo presentan problemas serios de credibilidad. Esto no hace a la credibilidad. La credibilidad como se define en el informe es “la confianza de los usuarios sobre el proceso del Registro Administrativo. Se debe considerar que al evaluar este criterio se evalúan políticas y prácticas objetivas para el diseño, recolección, almacenamiento, procesamiento y difusión de los datos.” Entonces tener una alta o baja tasa de no respuesta no implica tener una alta o baja credibilidad, sino en cómo se desarrollan las políticas y procedimientos en la entidad para lograr esa tasa de respuesta.

La tasa de no respuesta está relacionada con la cobertura (o defecto de cobertura, más precisamente). La diferencia es que el defecto de cobertura mide los casos que deberían estar en el RA y no lo están y la no respuesta contabiliza los casos que están en el registro pero con todas (o la gran mayoría) de las variables en blanco, salvo las variables de identificación que sí tienen datos (por eso no se contabilizan como defecto de cobertura, pues son casos válidos).

Supongamos una entidad que recoge la información del registro en formatos papel y una parte importante de estos (30%) llega en forma totalmente ilegible (salvo los datos identificatorios que están preimpresos), y no se han podido utilizar dispositivos móviles de captura (por diferentes razones), pero al momento de ingresar la información al sistema informático los digitadores entran lo que les parece o interpretan, logrando así una tasa de no respuesta de 0%. Podríamos decir, entonces, que esta entidad goza de una excelente credibilidad, cuando en realidad no es así. Por el contrario, si otra entidad en el mismo contexto tiene procedimientos claros y documentados, sus digitadores son conscientes de lo que ingresan al sistema (son capacitados en forma permanente), realiza validaciones y auditorías, hace pública toda su información, metodología, etc., pero como tiene una tasa de no respuesta del 30% entonces su credibilidad sería muy mala, cuando sabemos que es lo contrario.

Como ejemplo encontramos la Agencia Estatal de Vivienda de Bolivia, que en casi todos los criterios de credibilidad obtiene el máximo puntaje, excepto en porcentaje de no respuesta (calificación = 0, debido a que tiene más de 15% de no respuesta).

Por eso es que tal vez sea mejor considerar la tasa de no respuesta como parte de la dimensión de precisión y exactitud.

Además, la pregunta sobre la no respuesta no es clara, ha llevado a varias interpretaciones tal como se expresa en el informe de diagnóstico nacional de Bolivia. Donde se señala que algunos entrevistados interpretaron la no respuesta como falta de cobertura, otros la consideraron como variables sin dato y algunos la asociaron a casos del registro que no completan las condiciones exigidas. También en los informes de

diagnóstico de Ecuador y Colombia se señala que esta pregunta es incomprensible en el primer caso y ambigua en el segundo.

La pregunta 31. “Se cuenta con un sistema informático para la recolección del RA.” no es clara, pues no se indica si se refiere a un sistema para captura de datos de formatos papel o si son dispositivos móviles o una aplicación web. Por ejemplo, en el caso de la Agencia Estatal de Vivienda de Bolivia respondió 0 en esta pregunta, es decir, que no se cuenta con ningún sistema, lo cual hace dudar si se comprendió esta pregunta, pues las respuestas al resto de las preguntas sobre precisión y exactitud no son coherentes con ésta. En la pregunta 32 se respondió que sí se cuenta con un sistema informático para el procesamiento del RA, que se cuenta con una base de datos y recursos tecnológicos suficientes, etc. Tal vez se entendió que la pregunta estaba referida a si se cuenta con un sistema de captura en línea o formularios electrónicos.

Oportunidad y puntualidad

Este criterio debería ser de puntualidad solamente, pues entre sus indicadores (preguntas) no hay ninguno referido a oportunidad (lapso entre el período de referencia de los datos y la fecha de publicación de los primeros resultados estadísticos) o como se define en el CECRA: “tiempo transcurrido en el que se tiene disponible la información del Registro Administrativo y ésta aún es relevante para el seguimiento y evaluación o la toma de decisiones respecto al fenómeno que estudia o que es objeto de medición”

La pregunta 35 está orientada a los usuarios finales del producto estadístico.

35. “Los resultados del análisis estadístico del RA se difunden cumpliendo a tiempo con un calendario de difusión”

La pregunta 37 está orientada a los INE o unidad/departamento interno de la entidad que hace el análisis estadístico y/o genera el registro estadístico.

37. “Se tienen definidos calendarios de entrega de información con los usuarios del RA y se cumplen”

Se puede cumplir perfectamente con el calendario de difusión establecido (puntualidad) pero el lapso entre el período de referencia de los datos (en el momento de captura del RA o la fecha de referencia) y el momento en que se publican (de acuerdo al calendario preestablecido) no es lo que necesitan los usuarios (no es oportuno). Muchas veces es el criterio de calidad más importante para determinados usuarios.

Pertinencia o relevancia

La pregunta 41. “Se conocen y atienden plenamente las necesidades de información de los usuarios.” y su guía “Es importante asegurarse de que la información satisface las necesidades prioritarias, para lo cual se le produce. Esta pregunta evalúa el criterio de pertinencia.” Se refiere al cumplimiento de las necesidades de información de los usuarios, ¿cuáles? Se refiere a los usuarios del RA los INE o unidades/departamentos internos de estadística, o son usuarios finales de las estadísticas o por el contrario, son los usuarios administrativos del RA como tal, para lo cual fue creado el registro.

Como está dentro del capítulo “Análisis, entrega y difusión” podemos pensar que se trata de usuarios de los datos con fines estadísticos, pero no está claro si se trata de los

INE que utilizan el RA para generar un registro estadístico y estadísticas directamente a partir de éste o si se trata de usuarios finales que utilizan las estadísticas que genera la entidad a partir del RE creado a partir del/los RA(s).

Coherencia

Los indicadores que forman parte del criterio de coherencia permiten hacer una valoración muy parcial de esta dimensión de la calidad. Algo que hemos encontrado muy habitualmente en los RA es que utilizan clasificadores estandarizados pero en muchos casos los han incorporado muy recientemente y la mayor parte de la información del RA (histórica) fue ingresada sin aplicar dichos clasificadores, lo cual lleva a que los datos no sean comparables a lo largo del tiempo. Es así que dentro de la dimensión de coherencia se debería incluir como un sub-criterio la “comparabilidad” (ciertos modelos de evaluación de la calidad de las estadísticas lo consideran una dimensión como tal, Eurostat, Mercosur). En este sentido, bastaría con incluir una pregunta (o modificar las preguntas 13 y 13a) para indagar desde qué momento se utilizan los clasificadores estandarizados, u otra pregunta para saber desde cuándo la serie de tiempo es comparable.

Otras observaciones

El informe de diagnóstico nacional de Colombia incluye muy pocos RA en comparación con el resto de los países.

La matriz integrada de evaluación de la calidad de los RA de Ecuador sólo contiene calificaciones de RA de población. Asimismo, el informe de diagnóstico de inmuebles incluye las calificaciones de los indicadores de calidad del RA de catastro solamente.

Es necesario recabar la información faltante a los efectos de la evaluación de la situación actual y el grado de avance en cuanto al uso de RA con fines estadísticos y su integración en un sistema de registros estadísticos.

Cabe señalar que la herramienta HECRA²⁷ (Herramienta de Evaluación de la Calidad de los Registros Administrativos) del Banco Mundial toma como principal marco de referencia la experiencia de la Oficina de Estadística de Holanda. Si bien considera las experiencias del DANE de Colombia y de los países escandinavos, no está basada principalmente en estas metodologías y experiencias como se señala en el informe final de la consultoría de febrero de 2015.

Debido a ciertas inconsistencias o malinterpretaciones de las preguntas de los cuestionarios CECRA y CECAP descritas anteriormente, es necesario validar los resultados de los informes de diagnóstico nacional y las matrices integradas durante las visitas a los países.

En este punto se entiende necesario recabar información adicional sobre los RA evaluados y validar la información de los informes de diagnóstico nacionales, para lo

²⁷ Segui Stagno, Federico (2012). *Guía de la herramienta para la evaluación de la calidad de Registros Administrativos (HECRA) a ser usados con fines estadísticos*. Banco Mundial. México, 2012

cual se ha preparado la Ficha de requerimientos de información a cada país de acuerdo a los hallazgos encontrados (ver anexo I.2).

Anexo I.2 – Ficha de requerimientos de información a cada país

1. Por favor, describa brevemente los registros administrativos (RRAA) que son o podrían ser fuente de información del registro integrado de inmuebles.
 - a) ¿cuáles son las claves de identificación utilizadas para unir dichos registros? Si existen diferentes identificadores utilizados en distintos RRAA, por favor indique si existe una correspondencia entre éstos.
 - b) ¿cuál es la tasa de cobertura de cada RRAA?
 - c) ¿está previsto en la normativa vigente que el INE tiene acceso a los datos de las entidades responsables de estos RRAA? Por favor, indique para cada RRAA en qué caso sí y en cuáles no.
 - d) ¿el INE tiene acceso a las bases de datos o algún archivo con los datos de estos RRAA?
 - e) ¿el INE ya ha realizado intercambios de información (archivos) con las entidades responsables de estos RRAA? ¿con cuáles?
 - f) ¿cuál es la periodicidad de estas actualizaciones o intercambios?
 - g) en el caso de no tener aún intercambio de archivos con alguna entidad, por favor indique el motivo.
2. Registro catastral. Claves de identificación, domicilios (estandarización, actualización), ¿cuál es la clave o claves de identificación del registro? ¿cómo se actualizan los domicilios de cada predio o vivienda? ¿están estandarizados los domicilios, es decir, se utiliza una nomenclatura estándar? ¿cómo se vincula este registro con otros registros inmuebles y con las personas?
3. Describir el procedimiento de estandarización de domicilios que se está llevando a cabo o se piensa desarrollar.
4. ¿Cómo se hace la actualización de domicilios de las personas? Si no hay una obligación de actualizar el domicilio cuando se muda una persona o familia, ¿en qué momento se actualiza este cambio y quién lo hace (proveedores de energía eléctrica o agua, registros de educación o salud, programas de beneficiarios del estado, registro electoral, registro civil, etc)?
5. Autoridad Reguladora de Servicios de Energía Eléctrica (qué información contiene el RRAA, claves de identificación, variables, sobre todo domicilios, si está asociado al nro de predio o padrón o a la vivienda y si el contrato es con el propietario o el inquilino u ocupante o ambos).
6. Autoridad Reguladora de Servicios de Agua Potable (ídem anterior).
7. Por favor, describa brevemente los registros administrativos (RRAA) que son o podrían ser fuente de información del registro integrado de personas.
 - a) ¿cuáles son las claves de identificación utilizadas para unir dichos registros? Si existen diferentes identificadores utilizados en distintos RRAA, por favor indique si existe una correspondencia entre éstos.
 - b) ¿cuál es la tasa de cobertura de cada RRAA?
 - c) ¿está previsto en la normativa vigente que el INE tiene acceso a los datos de las entidades responsables de estos RRAA? Por favor, indique para cada RRAA en qué caso sí y en cuáles no.
 - d) ¿el INE tiene acceso a las bases de datos o algún archivo con los datos de estos RRAA?

- e) ¿el INE ya ha realizado intercambios de información (archivos) con las entidades responsables de estos RRAA? ¿con cuáles?
 - f) ¿cuál es la periodicidad de estas actualizaciones o intercambios?
 - g) en el caso de no tener aún intercambio de archivos con alguna entidad, por favor indique el motivo.
8. ¿Existe algún padrón centralizado de beneficiarios del estado, es decir, algún registro centralizado (bajo la responsabilidad del Min Desarrollo Social, o Min Planeación, o similar) de las personas que reciben algún tipo de beneficio del estado (no sólo personas en situación de vulnerabilidad, sino cualquier persona que reciba algún beneficio)? Por favor, describa este padrón, cómo se actualiza, periodicidad, fuentes que lo alimentan.
 9. ¿Existe algún procedimiento de caracterización de las personas/hogares en cuanto a su condición socio-económica a los efectos de priorizar la entrega de beneficios por parte del estado a las personas/hogares que están dentro de determinados puntajes o características? Si es así:
 - ¿cómo se realiza este procedimiento, se aplica una encuesta, auto-declaración, etc?
 - ¿existe una base de datos con esta información?
 - ¿cuándo y cómo se actualiza?
 - ¿el INE tiene acceso a esta base de datos?

Interoperabilidad de los RRAA. Variables clave a utilizar para unir los registros en un sistema integrado:

10. ¿Cómo se vinculan las personas con los inmuebles? Se utiliza una clave de identificación única? O se hace a través del domicilio de las personas?
 11. ¿Cuáles son los RRAA de salud y educación de las personas que se utilizan o se utilizarían y cómo se vinculan éstos con las personas (claves de identificación, principales variables)?
 12. ¿Cuáles creen que son las principales dificultades/debilidades con respecto a la integración o interoperabilidad de los RRAA de personas e inmuebles?
 13. ¿Cuáles son las principales fortalezas en cuanto a la integración o interoperabilidad de los RRAA de personas e inmuebles?
14. Aspectos de la calidad de los datos de los RRAA no contemplados en el cuestionario CECRA, pero necesarios para evaluar la utilidad de los RRAA con fines estadísticos:
 - a) Cobertura de los RRAA
 - b) tasa de respuesta de las principales variables de cada RRAA
 - c) si se realizan imputaciones ¿cuál es la tasa de imputación por variable?
 - d) ¿desde qué fecha se producen estadísticas a partir del RRAA en forma ininterrumpida? (comparabilidad en el tiempo, longitud de las series temporales)
 - e) tasa de errores de clasificación (para variables que usen clasificadores)
 - f) lapso entre el período de referencia de los datos y la fecha de publicación de los principales resultados estadísticos (oportunidad)
 - g) porcentaje de unidades o casos que pertenecen a la población objetivo creadas durante el año "t" que han sido registradas antes del fin del año t+1 (actualidad de los datos)

- h) tasa de registros (unidades, casos, filas) con clave única
- i) tasa de registros múltiples de la misma unidad

15. ¿Se han realizado pruebas o experiencias piloto en cuanto a la integración de RRAA para su uso con fines estadísticos (por ejemplo, experiencia de San Andrés en el caso de Colombia o Galápagos en el caso de Ecuador)? Por favor describa las principales conclusiones.