

# Hacia un sistema estadístico integrado y basado en registros





Hacia un sistema  
estadístico integrado  
y basado en registros

Copyright © 2021 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Nótese que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Banco Interamericano de Desarrollo  
1300 New York Avenue, N.W.  
Washington, D.C. 20577  
[www.iadb.org](http://www.iadb.org)

El Sector de Instituciones para el Desarrollo fue responsable de la producción de la publicación.

**Colaboradores externos:**

**Coordinación de la producción editorial:** Sarah Schineller (A&S Information Partners, LLC)

**Revisión editorial:** Clara Sarcone

**Diagramación:** Gastón Cleiman

# Hacia un sistema estadístico integrado y basado en registros

Preparado para el Banco Interamericano de Desarrollo por:  
Anders Wallgren y Britt Wallgren

## Resumen\*

Esta publicación describe la forma en que los países de América Latina y el Caribe pueden unirse a la revolución en los sistemas estadísticos, al pasar de la recopilación de datos basada en marcos geográficos a una con base en registros administrativos, y las ventajas de realizar ese cambio. Los países en el norte de Europa ya han pasado de un sistema estadístico tradicional basado en marcos muestrales de área a un sistema basado en registros, en el cual todas las encuestas se sustentan en registros estadísticos. Entre las ventajas clave del cambio se encuentran: (i) menores costos de producción, (ii) potencial para niveles más altos de desagregación geográfica y mayor frecuencia y (iii) reducción de la carga sobre los informantes al seguir la máxima de pregunta una vez, usa muchas veces. La evidencia de Colombia, Ecuador, México y Perú apunta hacia la viabilidad de esta transición en la región; sin embargo, para aprovechar mejor esta estrategia, los países deben invertir para mejorar la calidad y cobertura de sus sistemas administrativos y deben crear un sistema de registro integrado que permita el uso eficiente de los datos y asegure la consistencia y coherencia entre los registros estadísticos.

**Códigos JEL:** C1, C8, C19, H11

**Palabras clave:** estadística, registros estadísticos, registros administrativos, encuestas

\* Anders Wallgren y Britt Wallgren  
BA Statistiksystem AB, antes Statistics Sweden  
Correo electrónico: ba.statisti@telia.com

# Índice

<b>Prefacio</b>	<b>9</b>
<b>1 Conceptos básicos</b>	<b>11</b>
1.1 ¿Qué es una encuesta estadística?	<b>11</b>
1.2 ¿Qué es un registro administrativo?	<b>12</b>
1.3 ¿Qué es una encuesta por registro?	<b>13</b>
<b>2 ¿Por qué y cuándo deberían usarse registros administrativos?</b>	<b>17</b>
2.1 Dos modelos del sistema nacional de estadística	<b>17</b>
2.2 ¿Qué es un sistema integrado?	<b>18</b>
2.3 Comparación de costos y calidad	<b>23</b>
2.4 Combinar métodos: registros y muestreo de área	<b>25</b>
2.5 En resumen: por qué y cuándo usar los registros administrativos	<b>26</b>
<b>3 ¿Cómo se pueden utilizar los registros administrativos?</b>	<b>27</b>
3.1 Usar los números de identidad y proteger la privacidad	<b>27</b>
3.2 ¿Cómo utilizar los registros administrativos?	<b>32</b>
3.3 Usos de un registro administrativo	<b>33</b>
3.4 Modelo del proceso de producción y modelo del sistema de registros	<b>35</b>
<b>4 ¿Cómo debería hacerse la transición de los sistemas anteriores a los nuevos?</b>	<b>37</b>
4.1 Modelo de proceso de producción	<b>38</b>
4.2 Modelo de sistema de registros	<b>43</b>
4.3 Aseguramiento de calidad	<b>48</b>
<b>5 Problemas relacionados con el registro de población</b>	<b>50</b>
5.1 El registro civil y el sistema nacional de estadística	<b>51</b>
5.2 Cómo encontrar duplicados	<b>54</b>
5.3 Cobertura insuficiente y domicilios particulares incorrectos	<b>56</b>
<b>6 Conclusiones</b>	<b>64</b>
<b>Referencias</b>	<b>66</b>





# Prefacio

Contar con estadísticas oficiales oportunas y de calidad es fundamental para el diseño, implementación y evaluación de las políticas públicas en un país. Tradicionalmente, las estadísticas oficiales se han producido a partir de tres fuentes principales de información: censos, encuestas por muestreo y registros administrativos. De estas tres fuentes, la menos utilizada son los registros. En América Latina y el Caribe (ALC) las oficinas nacionales de estadística (ONE) basan la gran mayoría de su producción de datos en censos y encuestas. Sin embargo, este trabajo de campo requiere de una elevada inversión de recursos y tiempo y de una compleja planeación logística.

Inspirados por las experiencias exitosas de países como Canadá, Dinamarca, España, Noruega, República de Corea y Suecia, que operan bajo el principio de que el gobierno debe preguntar una sola vez y usar muchas veces a fin de reducir la carga tanto a los encuestadores como a los encuestados, cada vez más ONE de la región están empezando a explorar el enorme potencial de los registros administrativos como fuente de datos estadísticos. Esto cobra mayor sentido si se considera que los sistemas administrativos generan continuamente datos con gran nivel de desagregación, en rubros como registro civil (nacimientos, matrimonios, defunciones), educación, seguridad social, salud, empleo, medio ambiente, migración, turismo, etcétera.

Diversos factores han limitado el uso estadístico de los registros: falta de cobertura o estandarización de los mismos, insuficiente capacidad técnica o tecnológica para procesarlos e integrarlos, o marcos legales que dificultan su acceso por parte de las ONE. Para atender esta necesidad, desde hace varios años el Banco Interamericano de Desarrollo (BID) ha trabajado de cerca con los países de ALC y ha impulsado diversas iniciativas con miras a mejorar la calidad y la cobertura de los sistemas administrativos; fortalecer la capacidad de las ONE para que desempeñen un rol central en la coordinación, estandarización, intercambio e interoperabilidad de los registros; promover la creación de sistemas de registros integrados y coordinados; e instituir marcos legales estadísticos modernos que permitan un adecuado flujo y resguardo de la confidencialidad de los datos. Algunas de estas iniciativas impulsadas por el BID son: (i) la traducción al español y publicación del libro *Register-based Statistics: Administrative Data for Statistical Purposes* de Wallgren y Wallgren (2007), editado como *Estadísticas basadas en registros. Aprovechamiento estadístico de datos administrativos* (2012), en colaboración con el Instituto Nacional de Estadística y Geografía de México (INEGI) y que cuenta con una versión electrónica de 2016 (Wallgren y Wallgren, 2016); (ii) la realización de cursos de capacitación y asesoría técnica de los autores

del libro, Anders y Britt Wallgren, en Argentina, Belice, Barbados, Bolivia, Chile, Colombia, Costa Rica, Ecuador, El Salvador y México; (iii) la elaboración de una Ley Genérica sobre Estadísticas Oficiales para América Latina, en colaboración con la Comisión Económica para América Latina y el Caribe de las Naciones Unidas (CEPAL, 2020), para la modernización de los marcos legales, y (iv) la oferta del curso en línea Aprovechamiento Estadístico de los Registros Administrativos.

Gracias a estos y otros esfuerzos, algunos países de la región han dado pasos firmes en el tema. Por ejemplo: Brasil y México han logrado la vinculación de censos y encuestas de hogares con diferentes registros administrativos; Colombia y Costa Rica han incluido el número de identidad en el censo para facilitar la integración con otras fuentes administrativas de información socio-demográfica; Chile, Ecuador y Uruguay han hecho o tienen planeado hacer pruebas piloto del censo a partir de registros administrativos. Para seguir contribuyendo a estos avances, se espera que la presente publicación sirva como un referente conceptual y metodológico de utilidad para aquellos países que quieran incursionar o profundizar en el diseño, desarrollo e implementación de un sistema integrado de producción estadística basado en registros.

**Karla Yee Amézaga**

Consultora en capacidad estadística  
Banco Interamericano de Desarrollo

**José Antonio Mejía**

Especialista líder en modernización  
del Estado  
Banco Interamericano de Desarrollo

# 1 Conceptos básicos<sup>1</sup>

Dos principios constituyen la base de nuestras publicaciones sobre estadística basada en registros (Wallgren y Wallgren, 2012 y 2014): el enfoque de encuesta para datos administrativos y el enfoque de sistemas. El primero analiza estimados, estimadores y la calidad como en los libros sobre encuestas por muestreo. El enfoque de sistemas se construye alrededor del concepto de sistema de registros. Ambas publicaciones también analizan el sistema de producción de una oficina nacional de estadística (ONE) y el papel que desempeñan los registros administrativos en el diseño y desarrollo de dicho sistema. La presente publicación se basa en ese conocimiento.

Esta primera sección analiza los siguientes conceptos básicos: ¿qué es una encuesta estadística?, ¿qué es un registro?, y ¿qué es una encuesta por registro y cuál es la diferencia entre los registros administrativos y los estadísticos?

## 1.1 ¿Qué es una encuesta estadística?

La encuesta estadística es un término esencial utilizado por los estadísticos en todas las ONE. Sin embargo, para muchos de ellos el término encuesta es sinónimo de encuesta por muestreo. Esto puede causar confusión cuando se trata de una encuesta basada en registros administrativos. Para evitar dicha confusión, esta publicación utiliza la distinción entre los diferentes tipos de encuestas de la Oficina de Estadísticas de Canadá (2009). De esta manera, el término encuesta se usa genéricamente para abarcar cualquier actividad en la que se recopila o se obtienen datos estadísticos, la cual, de acuerdo con los Lineamientos de Calidad de la Oficina de Estadísticas de Canadá, incluye:

- un censo, que busca recopilar datos de todos los miembros de una población;
- una encuesta por muestreo, en la cual los datos se recopilan de una muestra (usualmente aleatoria) de los miembros de la población; y
- una encuesta por registro (que es el término empleado en esta publicación), en la que los datos se obtienen a partir de los registros que originalmente tienen fines no estadísticos.

De acuerdo con esta definición, una encuesta estadística es una actividad en la que se recopilan u obtienen datos estadísticos. Durante la encuesta se

<sup>1</sup> Esta sección se basa en la **sección 1.5** de Wallgren y Wallgren (2014).

genera una matriz de datos con microdatos estadísticos. Estos microdatos se utilizan finalmente para generar los estimados que son el objeto de la encuesta. En una encuesta por registro, se crean uno o más registros estadísticos a partir de registros administrativos.

## 1.2 ¿Qué es un registro administrativo?

Un registro administrativo se utiliza para almacenar registros sobre todos los objetos que se van a administrar y que el proceso administrativo requiere que puedan identificarse. La siguiente definición es válida para los registros administrativos y estadísticos:

- Un registro pretende ser una lista completa de los objetos en un grupo de objetos específicos o de población. Sin embargo, los datos en algunos objetos pueden faltar debido a deficiencias en la calidad.
- Los datos sobre la identidad de un objeto deberían estar disponibles con el fin de actualizar y expandir el registro con nuevos valores de variables para cada objeto.
- La lista completa e identidades conocidas son, por lo tanto, las características de un registro.
- Catálogo, directorio, lista, registro, padrón son diferentes términos para el mismo concepto. En esta publicación se usará el término registro.

Los siguientes son ejemplos de registros:

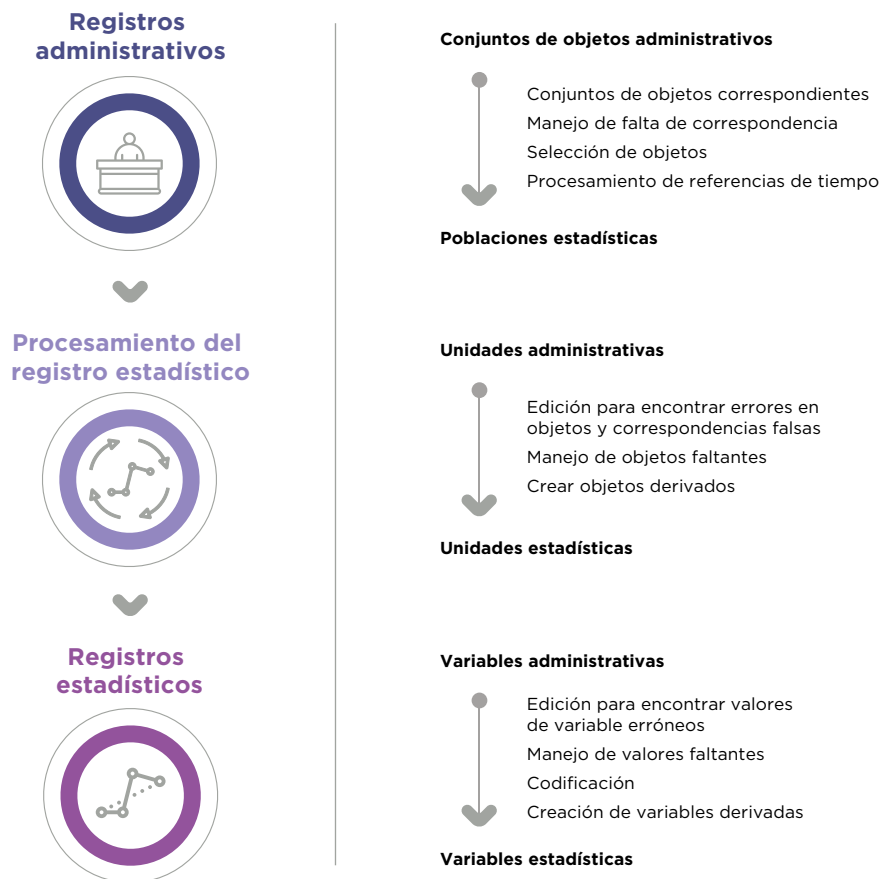
- El registro cívico, civil o nacional de población en un país produce registros de ciudadanía, nacimientos y fallecimientos.
- El autorregistro de ingresos de las personas aporta un registro de todos los contribuyentes en un año dado.
- En Suecia, las empresas con una facturación de 40 millones de coronas suecas o más deben entregar un informe mensual. Esto genera registros mensuales de todas las empresas que presentan dicho informe. Para empresas más pequeñas se obtienen registros trimestrales o anuales.
- Todas las transacciones de exportación e importación se registran en el organismo aduanero. Cada mes se generan registros con todas las transacciones de ese mes específico.
- Un archivo censal con datos sobre un censo de población y vivienda es un registro si contiene las identidades de las personas en el archivo.

### 1.3 ¿Qué es una encuesta por registro?

En la **sección 1.1** se explicó que en una encuesta por registro se crean uno o más registros estadísticos a partir de registros o archivos administrativos. Los datos administrativos originales se generan en sistemas administrativos públicos. Las definiciones de los conjuntos de objetos, objetos o unidades administrativas, así como las variables, se adaptan a los propósitos administrativos. Además, cada autoridad lleva a cabo procesos de control, corrección y de otro tipo que son adecuados para sus objetivos administrativos.

En general no se recomienda generar estadísticas directamente de los registros administrativos recibidos, ya que estos no se adaptan a los requisitos estadísticos. Los conjuntos y definiciones de los objetos y sus variables requieren un trabajo de edición. Como regla, siempre es necesario realizar una serie de pasos que constan de diferentes tipos de procesamientos de registros para que cumplan con los requisitos estadísticos para medir población, objetos o las unidades y variables estadísticas.

El **gráfico 1** muestra tres componentes importantes de este trabajo. Las personas tienden a usar los conceptos administrativos tal como están, pero si bien en algunos casos esto puede ser aceptable, en otros no. La manera en que se definen la población, unidades y variables de un registro estadístico son importantes para la calidad de las estadísticas que se producirán a partir de ese registro estadístico recién creado. El procesamiento estadístico de un registro, que busca transformar uno o varios registros administrativos en un registro estadístico, se describe en el **gráfico 1** como un proceso en el que se crean las poblaciones, unidades y variables estadísticas.

**Gráfico 1.** De los registros administrativos a los registros estadísticos

### Conjuntos de objetos administrativos: poblaciones estadísticas

Un registro de población estadístico se puede crear al combinar cinco fuentes: el registro civil para personas con números de identidad, el registro civil para personas sin números de identidad, el registro de nacimientos, el de fallecimientos y el de migración. La población estadística en un punto específico en el tiempo se puede crear usando referencias temporales, tales como fechas de nacimiento, fallecimientos, inmigración y emigración. Obsérvese que el registro civil administrativo y el registro de población estadístico creado por la oficina de estadística serían diferentes. El inventario de personas será distinto y las variables también; por ejemplo, si el domicilio particular en el registro estadístico se basa en otras fuentes además del registro civil, las variables de domicilio serán diferentes.

### Unidades administrativas: unidades estadísticas

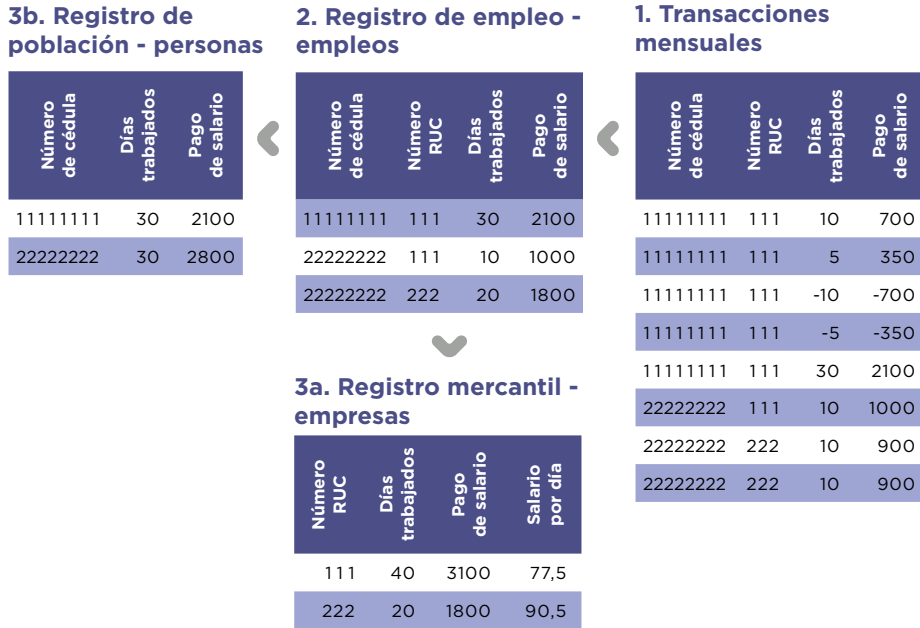
En el **gráfico 2** se usa el registro administrativo con las transacciones mensuales en un sistema de seguro social, para crear tres registros estadísticos para empleos, empresas y personas. Todos los empleadores en el sistema registran cada mes sus transacciones relativas a todos los empleados. En el **gráfico 2** se emplean dos tipos de números de identidad: los números de cédula que son los números de identificación personal (NIP) y los números del Registro Único de Contribuyentes (RUC) que corresponden al número de identidad empresarial (NIE).

En el registro 1 del **gráfico 2**, el empleador con identidad RUC 111 tiene dos empleados, con identidades de cédula 11111111 y 22222222. La persona 22222222 también es empleado del empleador con identidad RUC 222. En el sistema se han registrado todas las transacciones con correcciones y cambios. Cuando se entregó el registro administrativo 1 a la oficina estadística, se crearon los registros estadísticos 2, 3a y 3b. Las unidades administrativas “transacciones” en el registro 1 se transforman en las unidades estadísticas “empleos” en el registro 2, “empresas” en el registro 3a y “personas” en el registro 3b.

### Unidades administrativas: unidades estadísticas

Hay dos variables administrativas en el **gráfico 2**, “días trabajados por transacción” y “pago de salario por transacción”, en el registro administrativo 1 con transacciones mensuales. En los registros estadísticos 2, 3a y 3b se han derivado las variables estadísticas “días trabajados” y “pagos de salario” por empleos, empresas y personas. El registro 3a también tiene una variable estadística derivada “salario por día” para empresas.

En una encuesta por muestreo se definen las unidades estadísticas, los hogares o los establecimientos. Los entrevistadores revisan para garantizar que el tipo de unidad que se pretende estudiar sea la entrevistada. Las preguntas y definiciones en un cuestionario definen las variables estadísticas. En una encuesta por registro, se deben usar las unidades y las variables administrativas. A veces es posible agregar unidades administrativas y crear unidades estadísticas, como en el **gráfico 2**. Se pueden usar las variables administrativas para crear variables estadísticas derivadas, como en las matrices de datos 2, 3a y 3b del **gráfico 2**.

**Gráfico 2.** De unidades administrativas a unidades estadísticas



## 2 ¿Por qué y cuándo deberían usarse registros administrativos?

Cuando una ONE comienza a utilizar registros administrativos, se añade un nuevo factor de producción al sistema de producción. Las oportunidades que ofrecen los viejos métodos para llevar a cabo censos o encuestas por muestreo se mantienen; no obstante, en este nuevo modelo los registros administrativos también se utilizan con propósitos estadísticos, por lo que se suman nuevas oportunidades para el sistema de producción. En esta sección se analizan algunas consecuencias para el sistema de producción en el caso de que los registros administrativos estén disponibles para la producción estadística.

### 2.1 Dos modelos del sistema nacional de estadística

Aquí se comparan dos tipos diferentes de sistemas de producción nacionales: el sistema tradicional, en el que no se utilizan registros administrativos, y el sistema basado en registros, en el cual muchos registros estadísticos se crean con base en los registros de varias fuentes administrativas. En la actualidad muchos países se encuentran en algún punto entre estos dos extremos. Los registros administrativos se utilizan en cierta medida, pero el sistema nacional de producción todavía no se basa completamente en registros.

**Cuadro 1.** Comparación entre el sistema de producción tradicional y el basado en registros

<b>Sistema tradicional: sin registros</b>	<b>Sistema basado en registros: muchos registros</b>
<b>Censo</b> Basado en un marco muestral de área. Los entrevistadores tocan a la puerta y hacen entrevistas cara a cara.	<b>Censo</b> Un conjunto de registros estadísticos con datos anuales reemplaza al censo de población. Censo de explotación agrícola: se usa un registro para generar el marco muestral, se envían cuestionarios por correo a todos los participantes.
<b>Encuesta por muestreo</b> Basada en un marco muestral de área. Los entrevistadores tocan a la puerta de las unidades seleccionadas en una muestra y hacen entrevistas cara a cara.	<b>Encuesta por muestreo</b> Se usa un registro para crear el marco muestral, se envían los cuestionarios por correo, se llama por teléfono o se hacen entrevistas cara a cara a una muestra basada en registros (es decir, una muestra basada en el marco muestral de la lista).  <b>Encuesta por registro</b> Se crea un registro estadístico con registros administrativos.

El **cuadro 1** muestra que las principales diferencias metodológicas entre los dos sistemas de producción son los tipos de marcos y métodos de recopilación de datos. Los países sin un registro estadístico de población pueden clasificarse como países con un sistema tradicional; mientras que los países que llevan a cabo censos de población y vivienda basados completamente en registros se consideran países con un sistema de producción basado en registros.

Puesto que el registro de población es el registro estadístico más importante, la clasificación entre un sistema tradicional o uno basado en registros puede basarse en la existencia de un registro de población estadístico. Si la ONE ha sido capaz de crear un registro de población, será relativamente fácil crear otros registros.

## 2.2 ¿Qué es un sistema integrado?

La Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés) ha lanzado una estrategia global para el mejoramiento de las estadísticas agropecuarias y rurales en los países en desarrollo. La FAO (2010) analiza los sistemas integrados de estadística a nivel nacional. En el **recuadro 1** se incluyen algunas citas de este informe y se resalta en cursiva el texto fuertemente relacionado con lo que se entiende como un sistema integrado.

### **Recuadro 1. Estrategia global de la FAO para mejorar las estadísticas agropecuarias y rurales en los países en desarrollo**

En la integración y la metodología subyacente se consideran las dimensiones de calidad que incluyen la relevancia, la *exhaustividad*, la exactitud, la oportunidad, la accesibilidad, la *coherencia*, y la *comparabilidad*. El proceso para mejorar las estadísticas agropecuarias empezará con la integración de la agricultura en el sistema nacional de estadística. Esta integración se logrará mediante el desarrollo de un marco maestro de muestreo para la agricultura con el fin de asegurar la relevancia y la exhaustividad. Este marco se usará *para implementar un programa coordinado de recopilación de información para producir datos oportunos y exactos que sean coherentes y comparables*. Uno de los defectos de los sistemas estadísticos actuales, tanto en países industrializados como en países en desarrollo, es que los datos se recopilan por sector, usando diferentes encuestas y marcos de muestreo. Esta división de datos por sector no permite medir el impacto de una acción en un sector o en otro. En algunos casos, diferentes organizaciones producen estadísticas para los mismos ítems con resultados diferentes que confunden a los usuarios. Las encuestas se realizan con frecuencia de forma ad hoc sin vínculos a un marco de muestreo maestro o el uso de unidades georreferenciadas para la recopilación de datos. Por lo tanto, es difícil integrar los datos de diferentes encuestas para generar un análisis profundo con la tabulación cruzada de las variables. Los sistemas estadísticos integrados pueden resolver muchos de estos problemas, al evitar la duplicidad de esfuerzos, prevenir la publicación de estadísticas opuestas y asegurar el mejor uso de los recursos. Los conceptos, definiciones y clasificaciones se estandarizan, lo que permite la recopilación más sistemática de datos entre las diferentes fuentes.

**Fuente:** FAO (2010).

Las partes de un sistema integrado encajan perfectamente entre ellas. En el caso de un sistema estadístico, esto significa que las poblaciones y variables de distintas encuestas encajan entre sí. La exhaustividad, coherencia y comparabilidad son importantes factores de calidad. Los ministerios e institutos que llevan a cabo encuestas estadísticas deberán coordinar también la recopilación de datos para producir estadísticas sobre la sociedad. La FAO también quiere que los microdatos obtenidos de diferentes encuestas se combinen o integren a un nivel macro mediante la vinculación de registros, con el objetivo de lograr tabulaciones cruzadas de las variables. Según la FAO, los sistemas actuales tienen varias dificultades: la falta de cooperación entre los distintos actores lleva a la publicación de estadísticas en conflicto y a la duplicación de esfuerzos, lo que significa que no se están usando de manera eficiente los recursos económicos y humanos disponibles.

Wallgren y Wallgren (2014) analizan el concepto de sistema de registros. En el capítulo 1 definen un conjunto de principios para usar los registros administrativos con propósitos estadísticos. Los siguientes principios están relacionados con el sistema de registros:

- Principios del sistema: todos los registros estadísticos deberán incluirse en un sistema de registros coordinado. Este sistema garantizará que todos los datos se puedan integrar y usar de manera efectiva.
- Principio de consistencia: la consistencia de las poblaciones y las variables es necesaria para la coherencia de los estimados de las diferentes encuestas por registros.

En un sistema nacional de estadística con un sistema coordinado de registros es posible integrar todos los microdatos al sistema. En un sistema coordinado las poblaciones de los diferentes registros son consistentes debido a la cobertura, y las variables de los distintos registros no tienen valores contradictorios. Esto es lo que se entiende como un sistema estadístico integrado. Por supuesto que es difícil conseguir la consistencia perfecta, pero si los problemas de cobertura e inconsistencias en las variables son tolerables, puede afirmarse que el sistema es integrado. De esta manera, la integración de los microdatos será posible sin problemas perturbadores en lo que respecta a la falta de consistencia entre poblaciones y variables.

El **gráfico 3** ilustra un sistema de producción basado en registros. La base es el sistema de registros que consta de cuatro registros base (círculos) que definen las poblaciones en el sistema y cierto número de registros estadísticos con variables estadísticas importantes. Las líneas muestran los vínculos importantes entre las unidades de los diferentes registros. También se han añadido dos encuestas por muestreo: la de fuerza laboral y la de inversión.

**Gráfico 3.** Sistema de producción basado en registros (sistema de registros y encuestas por muestreo)

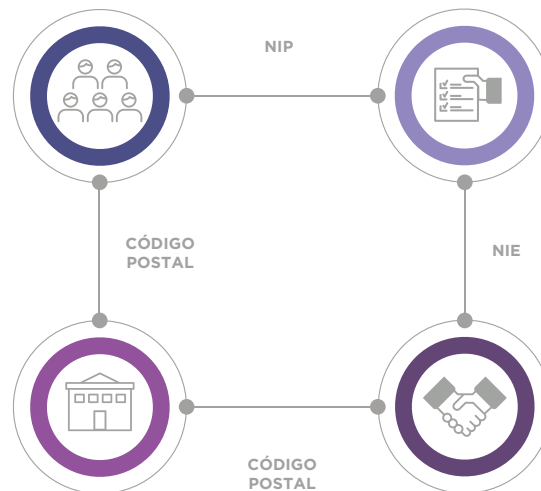
### Registro de población

#### Encuestas por muestreo de personas Encuestas de fuerza laboral

Registro de empleo  
Registro de educación  
Registro de ingresos y fiscalización  
Registro de ingresos longitudinal  
Registro de bienestar longitudinal  
Registros de investigación médica

### Registro de ocupación

Sistema de fiscalización preliminar  
Sistema de seguro social  
Educación primaria, alumnos  
Educación media superior, alumnos  
Registro de estudiantes universitarios



### Registro catastral

Base de datos geográfica  
(sistema de información geográfica [SIG])  
Cobertura del suelo, uso de suelo  
Registro de edificios  
Registro de viviendas

### Registro de empresas

Registro del impuesto al valor añadido (IVA)  
Registro de nóminas anuales  
Declaraciones anuales del impuesto  
sobre la renta  
Registro de comercio exterior  
Registro agropecuario

#### Encuestas por muestreo de empresas Encuesta de inversiones

**Nota:** NIP = número de identificación de personas; NIE = número de identificación de empresas.

Si las poblaciones y las variables tienen suficiente consistencia, el sistema en el **gráfico 3** es un sistema integrado. Este tipo de sistema tiene las siguientes características:

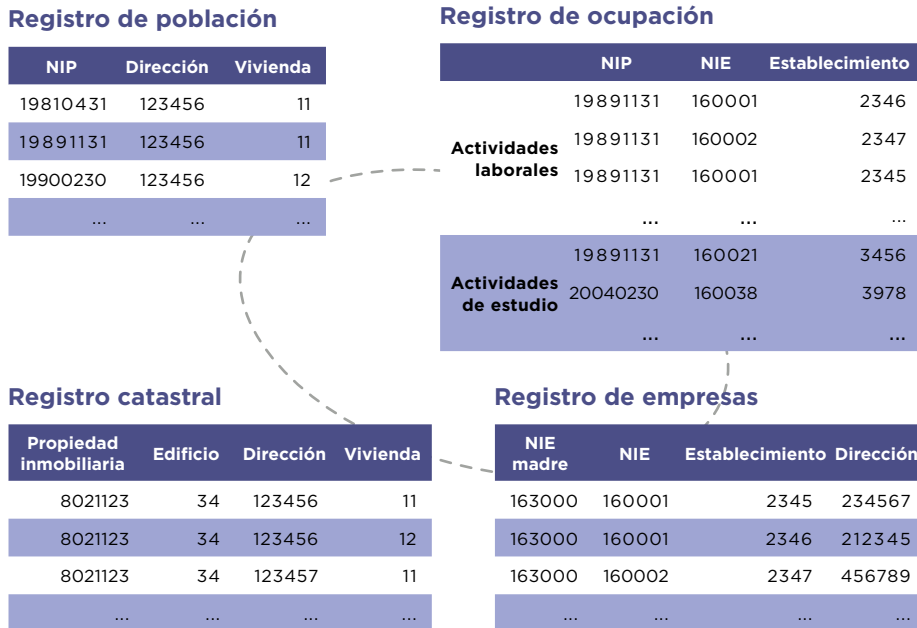
- Los microdatos de cualquier registro pueden combinarse con los microdatos de los demás registros. Por ejemplo, empleo, educación e ingresos se pueden combinar para crear un nuevo registro.
- Los microdatos de cualquier encuesta por muestreo se pueden combinar

con los microdatos de todos los registros. Por ejemplo, la encuesta de fuerza laboral puede usar variables del registro de empleo para mejorar los cálculos. También puede utilizar variables del registro de educación en lugar de hacer preguntas acerca de esta temática. Así mejorará la calidad y se reducirán los costos.

En contraste con la integralidad y flexibilidad de los datos de los registros, los microdatos de una encuesta por muestreo no se pueden combinar con los de otras encuestas por muestreo. Esto se debe a que, por regla, únicamente un pequeño número de unidades se incluye en más de una muestra, por lo que prácticamente no hay unidades en común que puedan combinarse. Esto explica por qué no se necesita el enfoque sistémico en un sistema de producción tradicional sin registros: es suficiente pensar en una encuesta por muestreo a la vez.

Por lo anterior, las estadísticas basadas en registros y las basadas en muestreo necesitan enfoques diferentes. El enfoque de sistema es fundamental para las estadísticas basadas en registros; en contraste, las basadas en muestreo no lo necesitan en absoluto. La bibliografía sobre teoría del muestreo y metodología de encuestas por muestreo se centran en discusiones sobre una encuesta por muestreo, no sobre un sistema de encuestas. Sin embargo, cuando se crean registros estadísticos se deben considerar simultáneamente varias fuentes a la vez, por lo que existen diferentes maneras de utilizar una misma fuente, así como distintas maneras de emplear las partes del sistema de registros.

El **gráfico 4** muestra las unidades estadísticas y sus números de identidad, en cuatro registros base. Con estas identidades, todos los registros se pueden vincular dentro del sistema. Puesto que los NIP o NIE se incluyen en encuestas por muestreo, estos también pueden vincularse al sistema. En el registro de ocupación se combinan los números de identidad NIP y NIE. La actividad es la unidad estadística, y la misma persona puede tener varias actividades de empleo y estudio durante un periodo específico. El registro de ocupación es el vínculo entre los registros de población y de empresas. Gracias a este vínculo, la estadística social y la económica son parte del mismo sistema de producción.

**Gráfico 4.** Los cuatro registros base y sus vínculos

Un sistema basado en registros también abre nuevas posibilidades para organizar el trabajo y eliminar la duplicación de esfuerzos. Mientras un equipo es responsable de mantener el registro de población y crear todas las poblaciones de personas en el sistema de producción, otro equipo es responsable de todas las variables de ingreso dentro del sistema, etc. Así, estas poblaciones y variables estandarizadas se pueden utilizar en todas las demás encuestas. Las estadísticas producidas serán más consistentes y coherentes en la medida que aumente la especialización y se profundicen las competencias.

Además, cuando el sistema de producción estadística se convierte en un sistema basado en registros, surgen nuevas posibilidades de evaluación de la calidad; por ejemplo, pueden detectarse problemas de cobertura al comparar las poblaciones de los registros relacionados. Los registros en el **gráfico 3** se pueden comparar de las siguientes maneras:

- ¿Todas las direcciones y viviendas del registro de población y las direcciones del registro empresarial están incluidas en el registro catastral?
- ¿Todas las personas del registro de ocupación están incluidas en el registro de población? ¿Todos los empleadores del registro de ocupación están contenidos en el registro de empresas?

Si se comparan variables similares en los distintos registros, y también variables de la encuesta por muestreo con variables similares en los registros, entonces se podrán encontrar los errores e inconsistencias en los valores de las variables. Un ejemplo importante de este tipo de trabajo es comparar la codificación de la actividad económica (Clasificación Internacional Industrial Uniforme [CIIU]) en distintas fuentes. A partir del registro de ocupación será posible encontrar a los empleadores de las personas entrevistadas en la encuesta de fuerza laboral. Entonces, la codificación de la actividad económica en la encuesta de fuerza laboral se puede comparar con la codificación en el registro de empresas.

En resumen, un sistema estadístico integrado basado en registros tiene las siguientes ventajas:

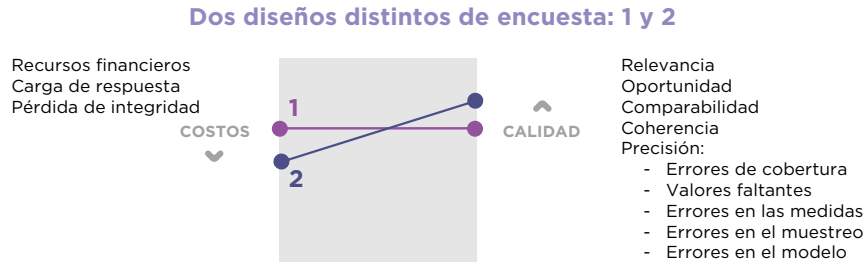
- Es posible integrar microdatos provenientes de distintas fuentes y crear nuevas combinaciones. En este sentido, los datos existentes se pueden usar de manera más eficiente.
- Se puede reducir la duplicación de trabajo y se puede mejorar la consistencia.
- Surgen nuevas maneras de realizar la evaluación de la calidad.

La creación de un sistema nacional integrado basado en registros requiere mucho trabajo y nuevas formas de pensar. Este proyecto de largo plazo necesita apoyo continuo de los administradores y políticos. La cooperación de los ministerios y autoridades administrativas también es necesaria para la implementación exitosa de nuevos métodos de producción estadística.

### 2.3 Comparación de costos y calidad

¿Por qué se deben usar los registros administrativos para fines estadísticos? Hay dos razones que explican este cambio en el sistema nacional de estadística: reducir los costos y mejorar la calidad. Cuando se trabaja en el diseño de encuestas, siempre se busca alcanzar estos dos objetivos.

En el **gráfico 5** se comparan dos diseños de encuestas. Si se asume que el diseño 1 es una encuesta de fuerza laboral en un sistema tradicional y el diseño 2 es una encuesta de fuerza laboral en un sistema basado en registros, entonces se pueden reducir los costos, ya que en el segundo sistema se pueden usar cuestionarios por correo o entrevistas por teléfono, en lugar de enviar entrevistadores a visitar una muestra de hogares del área. Además, la calidad puede mejorar cuando se reducen los errores de muestreo y de no respuesta al usar las variables auxiliares del sistema de registros.

**Gráfico 5.** Diseño de la encuesta: reducir costos y/o mejorar la calidad

Es necesario reducir los tres tipos de costos, por lo que se pueden ahorrar recursos financieros y disminuir la carga de la respuesta con la transición a un sistema basado en registros.

En muchos casos la recopilación de datos tradicional perturbará a los encuestados. Por ejemplo, cuando las personas atienden una llamada telefónica de una ONE durante la noche, pueden considerarla como una pérdida de privacidad. O el caso de una entrevista en la que el entrevistador puede haber obtenido información confidencial que podría ser una amenaza para la privacidad de los encuestados. Es posible reducir estos costos mediante las encuestas por registros.

En un sistema basado en registros se deben sustituir las identidades por números anónimos para proteger la integridad y el acceso a los microdatos en los registros debe estar restringido. Si no se toman estas medidas, los costos como las pérdidas de integridad pueden incrementarse.

Al analizar la calidad, se debe recordar que un sistema basado en registros abre nuevas oportunidades. Así, en lugar de un censo de población y vivienda tradicional y caro, se tiene un número de encuestas por registros que proporcionarán estadísticas anuales que no existían antes. Por ejemplo, las estadísticas anuales para pequeñas regiones solo serán posibles en un sistema de producción basado en registros o también podrán crearse registros longitudinales que resultarán muy importantes para los investigadores. En estos casos, se está comparando una situación sin datos a partir de encuestas importantes en el sistema tradicional con la calidad de las nuevas encuestas por registros.

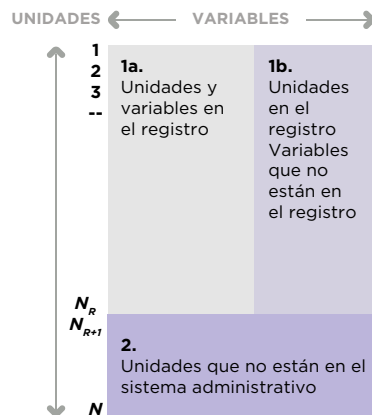


## 2.4 Combinar métodos: registros y muestreo de área

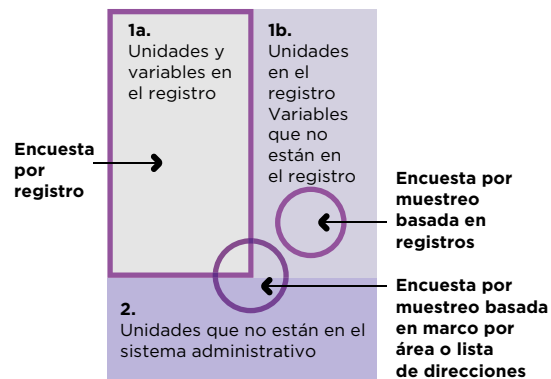
En un sistema basado en registros se necesitan tanto las encuestas por registros como las encuestas por muestreo como métodos complementarios. En Suecia se realizan encuestas por muestreo cuando no se pueden encontrar variables importantes en fuentes administrativas y también para estadísticas a corto plazo cuando los registros administrativos se generan demasiado tarde.

En varios países la cobertura de los sistemas administrativos no es perfecta. En los países en desarrollo pueden existir sectores informales, especialmente en regiones rurales, en los que las personas y las granjas no están incluidas en los registros administrativos. Por lo tanto, será necesario combinar encuestas basadas en registros con muestreo basado en áreas. Asimismo, algunos sistemas administrativos no se actualizan de manera regular, por lo que algunas variables importantes de los registros, como el domicilio, pueden ser incorrectas. Sin embargo, la magnitud de estos errores se puede medir con encuestas por muestreo basadas en áreas.

**Gráfico 6.**  
Diferentes partes de  
la población objetivo



**Gráfico 7.**  
Tres encuestas distintas



Cuando un registro estadístico tiene errores de cobertura debido a cobertura insuficiente, se puede combinar con una encuesta por muestreo basada en un marco de área. Dado que la muestra de área en teoría no tiene errores de cobertura, los estimados combinados pueden ser objetivos. De igual manera, una encuesta por muestreo basada en registros que utiliza como marco muestral un registro con cobertura insuficiente se puede combinar con un marco de área. En el **gráfico 6** la población objetivo está conformada por  $N$  unidades, y la población está dividida en dos partes: la parte 1 (1a y 1b) for-

mada por  $N_r$  unidades que se incluyen en el registro y la parte 2 formada por  $N - N_r$  unidades que conforman la cobertura insuficiente del registro.

La encuesta por registro en el **gráfico 7** se puede combinar con una muestra de área y la encuesta por muestreo basada en registros se puede combinar con una muestra por área. En el **apartado 5.3** se explica cómo se pueden usar estas muestras por área para evaluar la calidad y ajustar los errores de cobertura.

## 2.5 En resumen: por qué y cuándo usar los registros administrativos

Los sistemas administrativos del sector público generan enormes cantidades de datos que se almacenan en registros administrativos. Estos registros son una fuente que puede usarse para la producción estadística. Esta fuente está creciendo en volumen y, puesto que los sistemas administrativos mejoran gradualmente, la calidad de los datos administrativos que se generan por medio de estos sistemas irá siendo cada vez mejor. Usar los registros administrativos para la producción estadística reduce los costos y mejora la calidad. Los censos pueden remplazarse por encuestas por registros, y pueden desarrollarse nuevos tipos de estadísticas regionales y longitudinales. El sistema nacional puede transformarse en un sistema estadístico integrado con registros estadísticos consistentes y coherentes y encuestas por muestreo. Durante el periodo de transición quizá sea necesario combinar estadísticas basadas en registros con estimaciones basadas en muestras por área.

## 3 ¿Cómo se pueden utilizar los registros administrativos?

En Wallgren y Wallgren (2012; 2014) se analizan los métodos que deberían usarse en el trabajo con registros administrativos. Esta sección brinda una pequeña descripción de esos métodos y destaca aquellos que son nuevos o diferentes en un país que comienza a trabajar con registros. En esos países la ONE tuvo previamente acceso a las tablas estadísticas (macrodatos) producidas por los ministerios e institutos responsables de los registros administrativos. Sin embargo, un sistema estadístico integrado no se puede crear con macrodatos únicamente, sino que se debe basar en registros estadísticos coordinados. El primer cambio importante será que la ONE tenga acceso a todos los microdatos que se requieren para desarrollar y mantener el sistema nacional de registros estadísticos. Si se usan únicamente los macrodatos es posible comparar las fuentes y encontrar errores e inconsistencias; pero con los microdatos además es posible encontrar las causas de estos errores y hacer las correcciones para reducirlos.

### 3.1 Usar los números de identidad y proteger la privacidad

En un sistema tradicional basado en marcos de área, la ubicación de una unidad estadística es un factor que determina si la unidad se muestrea y entrevista. El nombre o identidad de una persona seleccionada en la muestra no es importante para la producción de estadísticas. Por consiguiente, la información sobre las identidades de las unidades no se usa en el proceso de producción. Los países con sistemas de producción basados en registros utilizan números de identidad en todas sus encuestas: censos, encuestas por muestreo y encuestas por registros. El registro de población y el de empresas se usan para crear marcos muestrales, y las unidades estadísticas se identifican con números únicos que se utilizan en todo el proceso de producción. En las encuestas por muestreo, cuando una muestra de personas/establecimientos ha sido seleccionada del marco muestral, las identidades se usan para contactar a las personas/establecimientos por correo o teléfono. En las encuestas por registros los microdatos tomados de los distintos registros se combinan para crear el conjunto de datos necesarios para la encuesta específica. Estas combinaciones se crean mediante vinculaciones de registros usando los números de identidad como claves para establecer la correspondencia.

En un sistema basado en registros tres aspectos son importantes en relación con las identidades:

1. La calidad de los números de identidad. Si la calidad es deficiente, deberá haber un proyecto a largo plazo para mejorar esta parte en el sistema nacional de registro. A corto plazo, los números de identidad se pueden complementar con nombres, fechas de nacimiento y lugares de nacimiento.
2. La vinculación de registros será una parte importante del proceso de producción cuando el sistema nacional de estadística se transforme en uno basado en registros. La razón es que todas las variables necesarias para una encuesta específica no aparecen en una sola fuente, sino que deben combinarse las fuentes.
3. El sistema de producción debe estar diseñado para proteger la privacidad usando números anónimos en lugar de números de identidad reales y garantizando el acceso restringido a las variables sensibles, como nombres y direcciones.

### Calidad de los números de identidad

Si un registro administrativo se usa de manera frecuente, la calidad de las identidades será buena. Los números de identidad nacional se utilizan para que la administración de los sistemas tributarios y de seguridad social sea eficiente y cumpla con los derechos de los ciudadanos. Estos sistemas administrativos regularmente sufren cambios y mejoras, y se desarrollan nuevos sistemas. Cuando un país comienza a usar datos administrativos con fines estadísticos, una tarea nueva e importante de los estadísticos es participar en el trabajo de desarrollo de los sistemas nacionales administrativos. Los estadísticos pueden contribuir con medidas de calidad generales y proporcionar información relacionada con las necesidades estadísticas para que los datos administrativos también puedan usarse en la producción estadística. Si el sistema nacional de registro incluye a todos los recién nacidos y nuevos inmigrantes, se evitan errores debido a cobertura insuficiente y duplicados. Los duplicados también pueden aparecer cuando las personas se registran más tarde en su vida (un registro anterior con un número de identidad previo pudo haberse olvidado).

### Vinculación de registros

En los países nórdicos la vinculación de los registros se ha practicado a gran escala durante varias décadas. Cada país tiene un sistema bien establecido de

números de identidad personal que se usa en todos los sistemas administrativos. Los números de identidad son de alta calidad y se utilizan en una gran cantidad de datos administrativos. Se emplea la vinculación determinista de registros, es decir, si los números de identidad coinciden, hay un vínculo; si no coinciden, no hay vínculo.

En un sistema basado en registros, siempre hay una vinculación al momento de crear el registro estadístico. Por ejemplo, si se combinan cuatro fuentes para crear un registro estadístico, entonces se realizan cuatro operaciones de vinculación. Para algunos registros se hacen más de 100 operaciones de vinculación, como sucede con los registros longitudinales que son muy importantes para la investigación. La conclusión de esto es clara: el índice de falta de correspondencia debe ser pequeño en cada operación de vinculación, de lo contrario, el registro final será difícil de usar.

En países sin buenos números de identidad en ocasiones se usa una vinculación de registros probabilística para combinar una encuesta por muestreo y un registro, o un censo y un registro. Para ello, se emplean combinaciones de claves de correspondencias: nombres, fechas de nacimiento y lugares de nacimiento en lugar de números de identidad. Algunas veces, estas variables aparecen registradas a mano en los formularios. Dado que frecuentemente los índices de falta de correspondencia son significativos, este método no es adecuado para un sistema estadístico basado en registros. En un sistema basado en registros las poblaciones e identidades se deben manejar para poder usar una vinculación de registros determinística.

En la Oficina Nacional de Estadística de Suecia el registro de población contiene únicamente a las personas registradas como residentes permanentes en Suecia. Todas estas personas tienen números de identidad válidos. Este principio genera cierta cobertura insuficiente, ya que hay personas que pertenecen a la población del censo deseado que no están formalmente registradas. No obstante, la ventaja es que la vinculación de los registros resulta muy fácil, ya que se pueden usar todos los registros administrativos que tienen números de identidad válidos.

Un país puede cambiar su sistema de números de identidad o tener sistemas paralelos, y se puede asignar números nuevos a algunas personas. La unidad que trabaja con el registro de población estadístico debe manejar esto. Se requiere un sistema capaz de vincular los números de identidad viejos con los nuevos de aquellas personas que hayan cambiado sus números de identidad. Además, los distintos tipos de números de identidad deberán transformarse en números de identidad anónimos que se usan en la oficina de estadística.

## Protección de la privacidad

Cuando un país cambia de un sistema nacional de estadística tradicional a uno basado en registros, se deben desarrollar nuevos métodos que protejan la privacidad. Las identidades pueden y deben manejarse de manera tal que en el nuevo sistema basado en registros la privacidad se proteja al menos tan bien como en el sistema tradicional.

En primer lugar se debe hacer un análisis de los datos administrativos y cómo se utiliza esta información en el proceso administrativo. Supóngase que el formulario de impuestos de un contribuyente X se usará para decidir sobre la renta gravable y el impuesto que debe pagar. Para revisar la información en el formulario de impuestos, los funcionarios que trabajan con la autoridad tributaria lo compararan con el formulario de impuestos del cónyuge de la persona X. Si estas personas también son dueñas de una empresa, el formulario de impuestos de la empresa también se debe revisar, ya que puede haber operaciones sospechosas entre X, el cónyuge y la empresa. Realizar una comparación con los formularios de impuestos del año anterior también es buena idea.

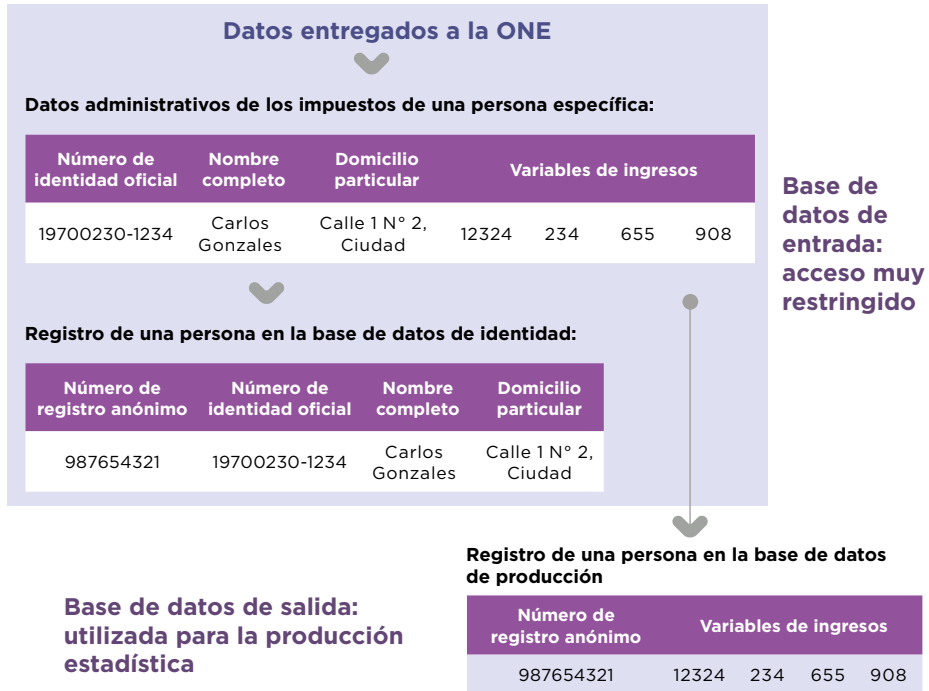
La situación descrita anteriormente es típica del trabajo con datos administrativos. Gracias a los sistemas de tecnologías de la información del sistema administrativo es fácil tener acceso a todos los datos pertinentes de una persona específica, en este caso se combinan seis formularios de impuestos que incluyen quizá 100 variables cada uno para el caso del contribuyente X. En conjunto con todos estos datos, se proporcionan los nombres y direcciones de personas y empleadores. A partir de la información disponible, los funcionarios en la oficina de impuestos deciden: ¿se puede confiar en el contribuyente X? Si la respuesta es sí, se determinan el ingreso gravable y el impuesto a pagar. Puesto que hay muchos funcionarios trabajando en revisar los formularios de impuestos, muchas personas en la oficina administrativa tienen acceso a los datos administrativos combinados con números de identidad oficial, nombres y direcciones. Estos funcionarios reconocen fácilmente a personas famosas o reconocidas.

El trabajo en la ONE se debe organizar de manera tal que el uso de números de identidad oficial, nombres y direcciones sea reducido y altamente restringido. De esta manera, la privacidad estará mejor protegida que antes. Este proceso se ilustra en el **gráfico 8**.

En un sistema nacional descentralizado las estadísticas se producen en los ministerios e institutos responsables de cada sistema administrativo. Estas estadísticas las generan funcionarios con acceso a los microdatos con nom-

bres y direcciones fácilmente reconocibles. Cuando el sistema nacional se vuelve más centralizado y basado en registros, esta producción se sustituye con el trabajo realizado en la ONE sin acceso a nombres y direcciones.

**Gráfico 8.** Un sistema informático que protege la identidad



Una unidad en la ONE, que consta de un grupo pequeño de personas, será la responsable de recibir todos los microdatos que se almacenan en una base de datos de entrada con acceso muy restringido. En esta unidad se eliminan los nombres y direcciones, y los números de identidad oficial se reemplazan por números anónimos. Una vez completado este trabajo, los datos anónimos se almacenan en la base de datos de salida y de producción, la cual pueden usarla para sus fines aquellos que necesitan acceso a los datos para la producción de estadísticas. De esta manera, la vinculación de los registros se lleva a cabo con los números anónimos en la base de datos de producción.

La información más sensible es el vínculo entre la identidad de la persona y su dirección. Por lo tanto, ha de imponerse la regla de que los registros estadísticos en la base de datos de salida solo usen información relacionada con el municipio o región estadística en la que viva la persona.

### 3.2 ¿Cómo utilizar los registros administrativos?

¿Cómo se usan los datos de encuestas por muestreo? Después de la fase de recopilación de datos hay disponible un conjunto de microdatos. Este conjunto de datos se usará de forma específica de acuerdo con el objeto de la encuesta. Una tradición bien establecida del diseño de encuestas se basa en estas premisas. Hay riesgo de pensar de manera similar cuando se quieren usar registros administrativos para la producción estadística: se tiene acceso a un registro administrativo y se trata de descubrir una manera de usar este conjunto de datos. Sin embargo, con esta manera tradicional de pensar no se obtendrá todo el potencial de las nuevas fuentes con datos administrativos.

Una encuesta por registro se define como el trabajo que se necesita para crear uno o más registros estadísticos con un fin específico. Los siguientes hechos explican por qué el diseño de una encuesta por registro es más complejo que el de una encuesta por muestreo:

- Un registro administrativo se puede usar para varias encuestas por registro y registros estadísticos.
- Muchos registros administrativos y estadísticos pueden necesitarse para una encuesta por registro.
- El registro estadístico puede ser usado por muchas otras personas para otras encuestas.

La encuesta de ingresos sueca basada en registros que se muestra en el **gráfico 9** ilustra lo anterior. Algunos registros administrativos externos se usan como datos de entrada. Además, se emplean algunos registros estadísticos internos obtenidos del sistema de registros de la Oficina Nacional de Estadísticas. Se crean tres registros estadísticos con estas fuentes, que constituyen los microdatos que se usan para la encuesta de ingresos. Algunas de las fuentes usadas como datos de entrada para la encuesta de ingresos también se emplean para otras encuestas por registro. Los formularios de impuestos obtenidos de la oficina tributaria nacional y las declaraciones de ingresos también se utilizan para generar estadísticas empresariales.



**Gráfico 9.** Encuesta de ingresos sueca basada en registros

### 3.3 Usos de un registro administrativo

Cuando la ONE tiene acceso a nuevas fuentes, ¿cómo debe usar los nuevos registros administrativos?, y ¿cómo se deben modificar las encuestas existentes? Para responder estas importantes preguntas, las fuentes administrativas deben ser analizadas por especialistas con experiencia en la materia y metodólogos con una buena visión general del sistema de producción estadística.

Un registro o fuente administrativos se pueden usar en cuatro formas diferentes. La primera, que implica usar el registro administrativo por sí solo, no aportará nada a un sistema integrado; pero las otras tres formas de usar un registro administrativo mejorarán el sistema de producción de manera que se vuelva más integrado.

### Método 1: por sí solo

Si la fuente tiene buena cobertura y sus variables son de buena calidad, la fuente puede utilizarse sola para producir estadísticas. Por ejemplo: en un sistema de registro de nacimientos (nacimiento), la identidad de la madre (número de identidad, nombre, dirección y fecha de nacimiento) y los datos médicos del bebé y su nacimiento son registrados por el personal del hospital. En este caso solo se tienen datos respecto de una parte de la población femenina, las madres recientes. Además de las estadísticas médicas, este registro se puede usar para producir estadísticas sobre el número de nacimientos por región y edad de la madre.

### Método 2: solo pero combinado con un registro base

El registro de población y el de empresas son dos registros base importantes que se usan para todas las encuestas relacionadas con personas o empresas en un sistema de producción basado en registros. Si un registro o fuente administrativos se combina con un registro base, se puede revisar y mejorar la calidad de este. De esta manera, será posible producir estadísticas consistentes y coherentes. El registro base contiene variables de clasificación importantes que pueden combinarse con el registro administrativo. Por ejemplo: si el registro de nacimientos y el de población pueden vincularse con la identidad de las madres, puede utilizarse el registro combinado para comparar a las madres recientes con el resto de las mujeres y producir estadísticas sobre el número de nacimientos por cada 1.000 mujeres por región y por edad de la madre.

### Método 3: combinado con un registro base y otros registros administrativos o estadísticos

En muchos casos un registro administrativo no tiene suficiente cobertura y/o el contenido de las variables está muy limitado. En consecuencia, la fuente no puede utilizarse por sí sola para la producción de estadísticas. Sin embargo, si se combinan varias fuentes, quizá sea posible utilizar el conjunto de datos combinados para estadísticas basadas en registros. Por ejemplo: si se pueden vincular el registro de nacimientos, el de población y el de educación estadístico con la identidad de las madres, este registro combinado se podría utilizar para producir estadísticas sobre el número de nacimientos por cada 1.000 mujeres por región, edad y nivel de educación de la madre.

#### Método 4: para mejorar otras encuestas (esto es, para mejorar el sistema de producción)

Si se comparan los registros, se podrán detectar diferencias relacionadas con la cobertura. Una vez que se hayan encontrado y entendido estos errores, también será posible corregirlos o reducirlos. Wallgren y Wallgren (2015a) analizan esos errores de cobertura y proponen cómo encontrar y reducir errores e inconsistencias entre las variables de clasificación que se encontraron al comparar la actividad económica y el sector en la Encuesta de Fuerza Laboral Sueca y la encuesta de Estadísticas Estructurales de Empresas. Por ejemplo: el registro de nacimientos antes mencionado puede utilizarse para mejorar la calidad del registro de población de tres maneras:

1. La información sobre la dirección de las madres recientes en el registro de nacimientos se puede usar para actualizar las direcciones en el registro de población.
2. Es posible que algunas madres no estén registradas en el registro de población administrativo. Esta cobertura insuficiente se puede reducir si los datos del registro de nacimientos de estas madres se añaden al registro de población estadístico.
3. Además, se puede incluir a los bebés recién nacidos en el registro de población estadístico.

Todas las encuestas por registro en países con sistemas estadísticos basados en registros (por ejemplo, los países nórdicos) utilizan un registro base como en los métodos 2 y 3. Los países con nuevos registros que aún no han desarrollado registros base de buena calidad empezarán con encuestas por registros de tipo simple, como las mencionadas en el método 1. Cuando ya se han desarrollado registros base, será posible crear encuestas por registros de acuerdo con los métodos 2 y 3. Hasta ahora, el método 4 no ha sido practicado ni siquiera en los países nórdicos, dado que comúnmente todos están muy ocupados con su propia encuesta y no tienen tiempo para realizar comparaciones con otras encuestas.

### 3.4 Modelo del proceso de producción y modelo del sistema de registros

El Modelo Genérico del Proceso de Producción Estadística (GSBPM, por sus siglas en inglés) ha sido analizado en algunas ONE. El foco de este modelo está en las distintas partes del proceso de producción para una encuesta. El **gráfico 10** ofrece una visión general del proceso y cada una de las siete partes que integran los muchos subprocesos. La idea principal es estandarizar los

métodos mediante listas de verificación, instrucciones y herramientas informáticas comunes.

**Gráfico 10.** Partes del proceso de producción de una encuesta



El modelo del proceso es aplicable a encuestas por muestreo y censos, pero no resulta de ayuda para el enfoque sistémico que es esencial en las encuestas por registro. El importante trabajo de crear, mantener y usar el sistema de registros no se incluye en el modelo del proceso. El modelo ilustra que el enfoque sistémico no se necesita en un sistema de producción tradicional sin registros, ya que es suficiente pensar en una encuesta por muestreo a la vez.

En contraste con la filosofía detrás del GSBPM, el modelo de sistema de registros tiene como objetivo estandarizar poblaciones y variables de manera que las encuestas sean consistentes y coherentes. Los datos se pueden usar de manera eficiente en un sistema integrado de registros, lo cual disminuye los costos y permite comparar poblaciones y variables, dando la oportunidad de encontrar y reducir errores. Los **gráficos 11 y 12** en el **apartado 4.1** ilustran los modelos de procesos que son adecuados para las encuestas por registro.

## 4 ¿Cómo debería hacerse la transición de los sistemas anteriores a los nuevos?

Un sistema estadístico tradicional basado en el muestreo de áreas puede ser más o menos integrado. La FAO (2010) señala que si todas las encuestas agropecuarias utilizan el mismo marco maestro de muestreo, y si las definiciones de variables siguen normas comunes, el sistema se volverá más integrado. Sin embargo, nunca será posible integrar microdatos que provengan de diferentes encuestas por muestreo, incluso si estas encuestas son consistentes y coherentes.

Cuando los datos tomados de registros administrativos se añaden al sistema, surgen nuevos problemas pero también nuevas posibilidades. Los datos del registro diferirán de los de la encuesta por muestreo, lo que significa que las inconsistencias son ahora aparentes. Además, diferentes registros tendrán distintas coberturas, valores de variables en conflicto y variables similares con diferentes definiciones. Sin embargo, no es posible combinar y comparar los microdatos que provienen de distintas fuentes. Se pueden encontrar errores e inconsistencias y se pueden realizar correcciones. De esta manera, se empezará a desarrollar gradualmente un sistema integrado de registros con poblaciones y variables consistentes. No obstante, esto requerirá mucho trabajo y deberán resolverse muchos problemas difíciles.

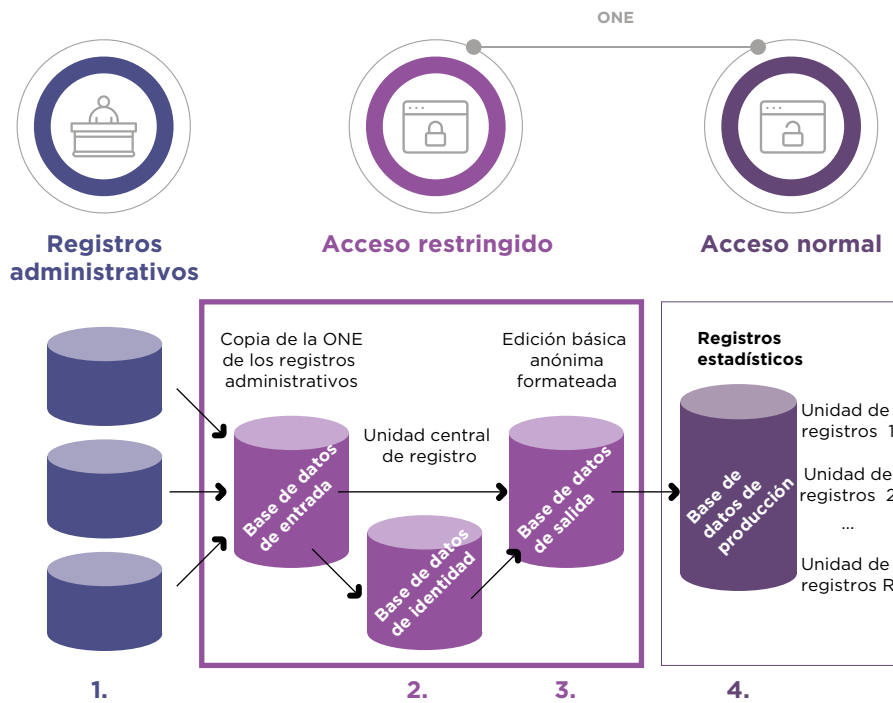
Para analizar cómo organizar el trabajo al tener acceso a los registros administrativos, se utilizan dos modelos diferentes. El modelo de proceso de producción<sup>2</sup> presentado en el **gráfico 10** se usará para describir cómo organizar el trabajo en la oficina de estadística. El modelo de sistema de registros (véase el **gráfico 3**) se usará para describir el uso de los registros administrativos para desarrollar un sistema integrado de registros estadísticos. Los problemas metodológicos relacionados con el proceso de producción se abordan con mayor detalle en la **sección 5**.

<sup>2</sup> Este modelo se desarrolló con David Chaín del Instituto Nacional de Estadística y Censos (INEC), durante una visita a Ecuador en 2015.

#### 4.1 Modelo de proceso de producción

El proceso de producción de registros administrativos a registros estadísticos se muestra en el **gráfico 1**, donde se destacan los métodos estadísticos. El mismo proceso se ilustra en el **gráfico 11**, que presenta la organización del trabajo. El proceso se estructura en cuatro pasos numerados del 1 al 4 en la parte inferior del gráfico.

**Gráfico 11.** Proceso de producción: transformar registros administrativos en registros estadísticos



- Paso 1: las autoridades administrativas entregan los registros administrativos a la ONE. La selección de unidades y variables que se entregarán debe corresponder a los requisitos estadísticos. La cooperación entre las autoridades administrativas y la ONE deberá estar bien definida y ser permanente. Para la actualización del registro de población y el de empresas, se deberán entregar únicamente los registros nuevos y los cambiados.
- Paso 2: con la base de datos de identidad, las variables identificadoras, como número de identidad personal, otros tipos de números de identidad y nombres, se sustituyen con números de registro anónimos. Únicamente se transmiten estos números de registro a la base de datos de

salida. Los registros con distintos números de identidad personal para la misma persona se vinculan con el mismo número de registro anónimo.

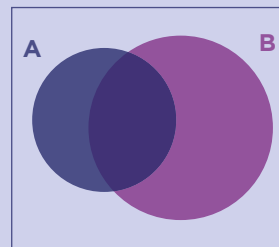
- Paso 3: las variables en la base de datos de entrada se transforman y formatean para que todas las variables de la base de datos de salida cumplan con las normas y formatos usados en la ONE. Además, se realiza una validación básica de las variables en cada registro administrativo en la base de datos de salida.
- Paso 4: el trabajo central de la creación de registros estadísticos se realiza en la base de datos de producción. Se combinan diferentes fuentes mediante la vinculación de registros, se crean poblaciones de registros, se crean unidades derivadas y variables derivadas, y se lleva a cabo la edición para consistencia.

### Contactos con las autoridades administrativas

Los contactos con las autoridades administrativas son necesarios para la ONE por diversos motivos. La ONE no solo necesita los microdatos, sino también los metadatos con definiciones de todas las variables administrativas. Además, se debe entender el proceso administrativo que genera los registros administrativos: ¿cómo se genera el conjunto de objetos administrativos? Estos conjuntos de objetos se usarán para crear poblaciones estadísticas y, por ese motivo, los estadísticos deben entender el sistema administrativo para poder evaluar la pertinencia de los datos.

Los contactos con la ONE también son importantes para las autoridades administrativas. Las autoridades podrán obtener nueva información de calidad. Se pueden realizar nuevos tipos de comparaciones con otras fuentes mediante el sistema de registros estadísticos. La falta de calidad estadística podrá indicar también falta de calidad administrativa. La cobertura insuficiente no es solo un problema estadístico, ya que de igual manera afecta el proceso administrativo. Las duplicaciones o domicilios particulares erróneos también son problemas administrativos.

La autoridad administrativa a menudo tiene muchos más datos de los que esperan los estadísticos. El rectángulo a la derecha representa los datos que tiene la autoridad administrativa. El área A representa los datos que la autoridad piensa que requieren los estadísticos y el área B, los datos que debe obtener la ONE. Para evitar malos entendidos, la selección de los datos que se enviarán deberá decidirse mediante conversaciones profundas entre ambas partes.



Los grupos de trabajo de la ONE deben manejar la cooperación con las autoridades administrativas. Estos grupos de trabajo deben estar formados por personal con competencia en la materia, en métodos estadísticos por registros y en tecnologías de la información. La cooperación entre estadísticos y los distintos ministerios y autoridades administrativas deberá ser un proyecto a largo plazo, con el objetivo de desarrollar sistemas nacionales estadísticos y administrativos. El **apartado 5.1** analiza el registro civil y el sistema estadístico.

### La unidad central de registro

La base de datos de identidad ha sido desarrollada con dos fines: proteger la integridad y facilitar el procesamiento de los registros. La integridad estará protegida ya que los nombres y números de identidad personal serán eliminados. El personal que trabaja con las estadísticas basadas en registros únicamente tendrá acceso a números de registro anónimos al generar los registros estadísticos.

Sin embargo, puede haber sistemas paralelos de números de identidad. La base de datos de identidad hará más fácil el proceso de producción con lo que se reducirá el riesgo de errores. Esto se debe a que todos los datos utilizados en la base de datos de producción se han editado de forma tal que las variables de identificación complicadas, como nombres y diferentes tipos de números de identidad, han sido sustituidas por un sistema unificado de números de registro fáciles de utilizar.

Se debe organizar una unidad central con un número limitado de personal, la cual será responsable de recibir todos los registros administrativos. Únicamente esta unidad deberá tener acceso a los servidores y bases de datos en donde se almacenen las copias de los registros administrativos.

La unidad central de registro también es responsable de sustituir los nombres y números de identidad oficiales por números de registro anónimos. Por lo tanto, únicamente la unidad central tendrá acceso a la base de datos de identidad en la cual están almacenados los enlaces entre identidades y números de registro anónimos.

La unidad central también será responsable del manejo de las identidades utilizadas en las encuestas por muestreo basadas en registros, las cuales se deben llevar a cabo de la siguiente manera:

- Se toma una muestra de personas al azar de un marco muestral creado con el registro de población. Como resultado se obtiene una lista con los números de registro anónimos para la muestra.



- Esta lista se envía a la unidad central, que reemplaza los números de registro con los nombres y direcciones utilizados para enviar los cuestionarios o proporciona los nombres y direcciones a los encuestadores. Las combinaciones de nombres/identidades reales y números de registro son la parte más confidencial del sistema y nunca deben existir fuera de la base de datos de identidad.
- Los datos recopilados se entregan a la unidad central, la cual elimina los nombres y direcciones y los reemplaza con los números de registro anónimos.
- El conjunto de datos anónimos se entrega a la unidad responsable de la encuesta por muestreo.

Además de manejar las identidades, la unidad de registro central también es responsable de dos tareas importantes en el procesamiento de los registros administrativos:

- Transformar todos los datos administrativos a los formatos y normas que se deben utilizar en la oficina de estadística. Si, por ejemplo, tres fuentes administrativas diferentes en la base de datos de entrada definen sexo en diferentes formas (M/F, 0/1 o 1/2 para hombres/mujeres) entonces en la base de datos de salida todos estos registros deben usar los mismos códigos (por ejemplo, 1 para hombres y 2 para mujeres). Este tipo de estandarización será mucho más importante cuando se cambie el sistema de producción basado en un marco muestral de área por uno basado en registros. En el sistema basado en registros se pueden combinar los microdatos de diferentes registros y encuestas. Por ese motivo, se deben usar los mismos formatos y normas en todos los registros y encuestas.
- Hacer una edición básica de cada registro administrativo. Esta edición consiste en dos tareas: buscar duplicados y encontrar y corregir valores erróneos de variables. Esto se describe en Wallgren y Wallgren (2014: **apartado 9.1.1**). El manejo de duplicados también se comenta en el **apartado 5.2** más adelante.

### ¡Hacerlo bien desde el principio!

Las bases de datos de identidad y las normas y formatos a utilizar para las variables del sistema se deben establecer al principio del proceso de modernización del sistema estadístico. Si no se sigue esta regla, deberán desarrollarse sistemas informáticos costosos que utilicen las identidades y formatos originales en las fuentes administrativas. Cuando las personas entiendan más adelante que la privacidad merece mayor protección y que la estandarización es necesaria, cambiar estos sistemas informáticos será muy costoso.

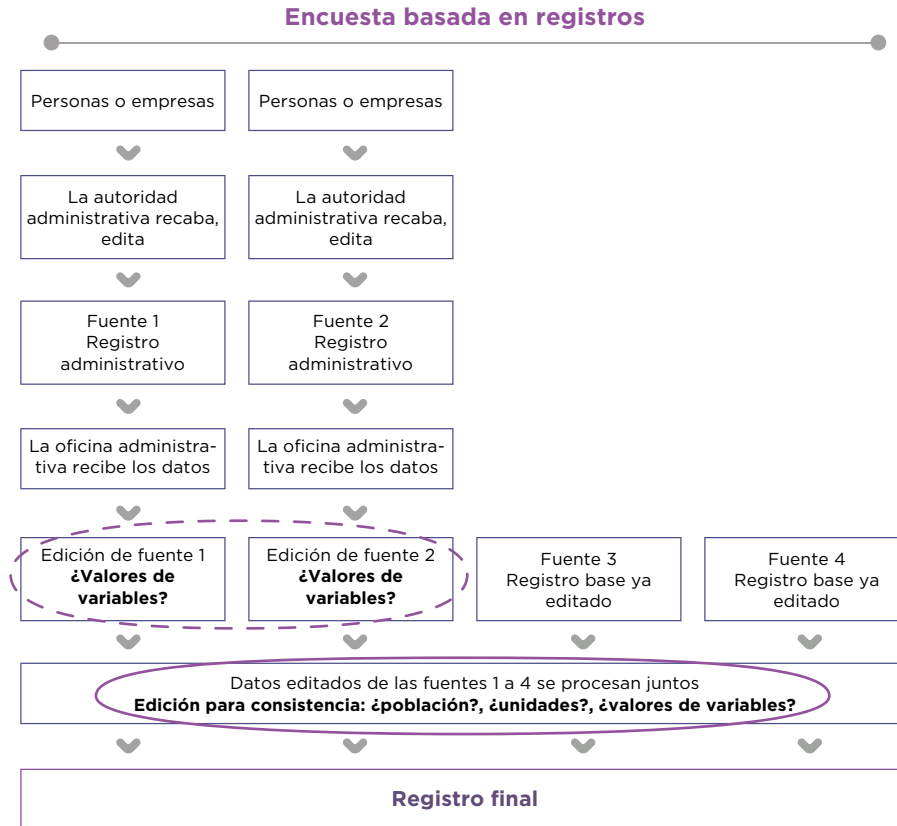
## La base de datos de producción

Diferentes unidades de registro en la base de datos de producción generan registros estadísticos. Cada unidad puede importar los registros administrativos que necesita desde la base de datos de salida. También puede utilizar los registros estadísticos del sistema (todos están almacenados en la base de datos de producción). La única identidad de las personas es el número de registro anónimo que se utiliza para vincular los registros. Las unidades de registro responsables crean poblaciones, unidades y variables estadísticas, las cuales se verifican con edición de consistencia, lo que permite descubrir y reducir los errores e inconsistencias.

Como se muestra en el **gráfico 12**, existen dos tipos de edición en la creación de registros estadísticos. La primera (que se marca como una elipse punteada) la realiza la unidad central de registro en la base de datos de salida cuando edita allí cada registro administrativo. Después de buscar los duplicados, se realiza la edición con los mismos métodos utilizados para editar encuestas por muestreo. El objetivo es encontrar y corregir los valores de variables erróneos.

El segundo tipo de edición (que figura como una elipse continua) es la edición de consistencia que llevan a cabo las unidades responsables de cada registro estadístico en la base de datos de producción, dado que se encontrarán muchos tipos de errores e inconsistencias diferentes al comparar las distintas fuentes. Además de encontrar valores de variables erróneos, también se hallarán errores de cobertura y errores relacionados con las unidades estadísticas.

El trabajo de edición que se realiza en las bases de datos de salida y de producción es la parte central del trabajo de aseguramiento de calidad de las estadísticas basadas en registros. La documentación de los resultados del trabajo de edición es importante ya que permitirá supervisar los cambios en la calidad de los datos administrativos en el tiempo. El aseguramiento de calidad se analiza en el **apartado 4.3**.

**Gráfico 12.** Dos tipos de edición en la creación de registros estadísticos

## 4.2 Modelo de sistema de registros

Los registros estadísticos se crean en la base de datos de producción. El modelo de sistema de registros se usa para analizar los siguientes temas:

- Cómo organizar el trabajo en relación con la competencia en la materia.
- Cómo organizar el trabajo para lograr consistencia.
- Cómo llevar a cabo la vinculación de registros.
- Cómo utilizar el sistema para diseño de encuestas.

### Competencia en la materia

Como se mostró previamente en el **gráfico 3**, el modelo de sistema de registros está estructurado por tipo de unidad estadística: personas, actividades, empresas y bienes inmuebles o terrenos. Esta estructura se relaciona con diferentes

campos de la materia. Por ejemplo, las estadísticas económicas se vinculan con el registro de empresas. Sin embargo, la relación es bastante débil, ya que las estadísticas económicas necesitan datos de las cuatro partes del sistema. Además, las estadísticas de personas requieren datos de todo el sistema.

¿Cómo organizar el trabajo en relación con la competencia en la materia? Considérense los cuatro registros con datos sobre personas como ejemplo: los registros de población, empleo, educación, ingreso e impuestos. Los registros anuales se crean cada año: la población de personas al 31 de diciembre, y la situación de empleo, educación e ingresos para la población de personas a esa fecha. Estos datos se pueden estructurar en cuatro registros diferentes o en un registro integral de estadísticas sociales (**cuadro 2**):

**Cuadro 2.** ¿Cuatro registros o un registro integral?

Registro de población	Registro de empleo	Registro de educación	Registro de ingresos e impuestos	Registro de estadísticas sociales
NIP	NIP	NIP	NIP	NIP
Edad	Edad	Edad	Edad	Edad
Sexo	Sexo	Sexo	Sexo	Sexo
Municipio donde vive	Municipio donde vive	Municipio donde vive	Municipio donde vive	Municipio donde vive
	Situación laboral	Nivel más alto de educación	Ingresos por empleo	Situación laboral
	CIU	Tipo de educación	Ingresos por ...	CIU
	Municipio donde trabaja		Ingresos disponibles	Municipio donde trabaja
				Nivel más alto de educación
				Tipo de educación
				Ingresos por empleo
				Ingresos por ...
				Ingresos disponibles

A partir del **cuadro 2** surgen dos preguntas: ¿cómo se debe organizar el trabajo?, y ¿cómo se deben almacenar los datos? La respuesta a la primera pregunta es que el trabajo se debe organizar para fortalecer el papel de la competencia en la materia. En el ejemplo debe haber cuatro unidades de registro con responsabilidades bien definidas.

- La unidad de encuestas de población es responsable de las estadísticas de población y del registro de población.

- La unidad de encuestas de empleo es responsable de las estadísticas del mercado laboral, de los registros de empleo y de seguro social.
- La unidad de encuestas de educación es responsable de las estadísticas de educación, educación primaria, educación media superior y registros universitarios.
- La unidad de encuestas de ingresos es responsable de las estadísticas de ingresos, el registro de ingresos e impuestos y el registro de ingresos longitudinales.

Cada una de estas unidades de registro requiere de diferentes competencias en la materia, los datos son entregados por distintas autoridades administrativas y tienen diversos tipos de usuarios. Por lo tanto, es preferible organizar cuatro unidades diferentes de registro cuando se haya establecido la producción de estadísticas.

La respuesta a la segunda pregunta es que los datos se pueden almacenar en cuatro registros independientes, con cierta redundancia, o se pueden almacenar en un registro integral sin redundancia. Con base en la experiencia de la Oficina Nacional de Estadísticas de Suecia, se recomienda la organización más sencilla con cuatro registros independientes en donde cada unidad de registro es responsable de su propia base de datos. La redundancia será insignificante debido a que los registros de empleo, educación e ingresos solamente importarán un número pequeño de variables del registro de población.

En este apartado se destaca la importancia de la competencia temática. Cuando el sistema estadístico se basa en los registros, el papel de este tipo de competencia cambia. Los contactos con las autoridades administrativas y la comprensión de los sistemas administrativos nacionales son ámbitos nuevos e importantes en los que se requiere esta competencia. ¿Cómo deben diseñarse los nuevos registros estadísticos para que proporcionen información importante sobre la sociedad? Se necesita competencia en la materia

El sistema de registros está estructurado por poblaciones y variables, es decir, por el contenido de los registros. Esto significa que las diferentes partes del sistema de registros están vinculadas a distintos tipos de competencia temática. El trabajo con los registros estadísticos se debe organizar para promover competencia en esta materia.

Durante la década de 1990, se crearon muchos registros estadísticos nuevos en la Oficina Nacional de Estadística de Suecia y los especialistas en la materia encabezaban el trabajo. Las unidades de registro podían contactar a las autoridades y futuros usuarios de las nuevas estadísticas y utilizar su creatividad para desarrollar nuevos registros. Se trataba de una organización

descentralizada en la cual los emprendedores tenían la libertad para crear nuevas estadísticas.

### Registros consistentes

En un sistema de producción perfectamente integrado, los diferentes registros y encuestas son consistentes en cuanto a cobertura y variables. El modelo de sistema de registros puede utilizarse para ilustrar las condiciones que se requieren para lograr la consistencia. Se debe organizar el trabajo para alcanzar esta consistencia y se debe involucrar a los administradores. Cuando se encuentran errores e inconsistencias se deben tomar decisiones: qué registros se deben corregir o cambiar y cómo se deben asignar los costos para las mejoras.

Las estadísticas basadas en registros necesitan del trabajo en equipo entre las diferentes unidades en relación con las poblaciones y las variables. Si una unidad trabaja con estadísticas mensuales y otra con estadísticas anuales, es posible olvidar esta coordinación debido a la diferencia en los periodos de referencia. Supervisar y fortalecer este trabajo en equipo será una tarea importante para los administradores. A continuación se proporcionan dos ejemplos de este tipo de cooperación: cuatro registros anuales y un registro de empleo.

**Cuatro registros anuales.** Primero, la unidad responsable del registro de población crea la población al 31 de diciembre. De ahí en adelante, las unidades responsables de los otros registros toman este registro de población con clasificaciones básicas por sexo, edad, región de residencia, etc. y agregan sus propias variables, como empleo, educación e ingresos (véase el **gráfico 13**). De esta manera, se logra una coordinación perfecta con respecto a poblaciones y variables.

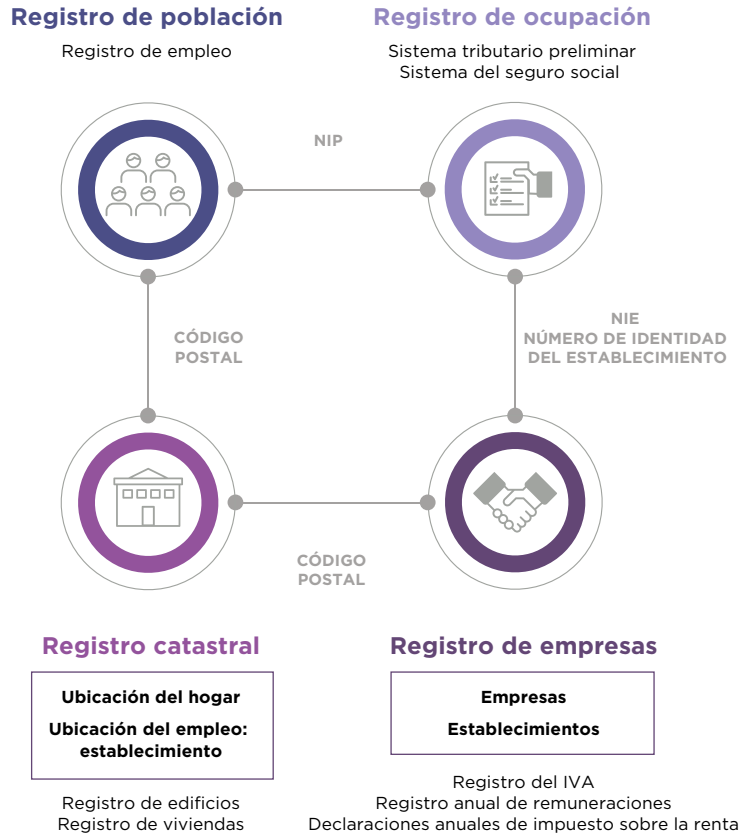
**Gráfico 13.** Cuatro registros anuales



#### Registro de población

Registro de empleo  
Registro de educación  
Registro de ingresos

**Registro de empleo.** El registro requiere cooperación con políticos y entre diferentes departamentos en la ONE, como se muestra en el **gráfico 14**.

**Gráfico 14.** Registro de empleo

En primer lugar, los empleadores con más de un establecimiento deben informar el lugar de trabajo de cada empleado. Puede ser necesaria una decisión política para que esta información forme parte del sistema de seguro social o fiscal.

En segundo lugar, se necesitan datos de todas las partes del sistema para crear un registro de empleo. Se utilizarán cinco identidades para la vinculación del registro: números de identidad de personas, empresas y establecimientos, y códigos postales para la ubicación de los hogares y establecimientos. La unidad responsable del registro de empresas debe crear un registro de todos los establecimientos, cuyo objetivo principal es respaldar el registro de empleo. Se necesita cooperación para lograr consistencia en poblaciones: la población de empleados y de personas autoempleadas debe ser consistente con el registro de población y la población de empleadores y personas autoempleadas debe ser consistente con el registro de empresas.

## Vinculación de registros y diseño de encuestas

Cuando se analiza la vinculación de registros en las publicaciones científicas, generalmente se hace referencia a la vinculación de un conjunto de datos para personas con otro conjunto de datos para personas. El modelo de sistema de registros muestra que en un sistema basado en registros la vinculación de registros está más avanzada y proporciona más posibilidades. En un sistema de este tipo se pueden vincular las personas tanto con empresas como con viviendas, y las actividades con personas o empresas.

Esta característica de un sistema integrado crea nuevas oportunidades para el diseño de encuestas. El registro de empleo descrito en las secciones anteriores es un ejemplo en el cual se combinan los datos de varias partes del sistema de registro. Otro ejemplo es el registro longitudinal para investigación creado por la Oficina Nacional de Estadística de Dinamarca (Stender, Thorsen y Andersen, 2015): un gran volumen de datos sobre actividades de trabajo y estudio, y la participación en diferentes programas de bienestar social se utiliza para describir el apego de la población al mercado laboral.

Se requiere competencia en la materia para crear registros de este tipo, y es importante trabajar junto con los investigadores académicos. Este análisis de la materia define el contenido deseado. Posteriormente, es necesario un procesamiento avanzado de los registros de muchas fuentes para crear el registro deseado.

### 4.3 Aseguramiento de calidad

El registro de población estadístico es la parte más importante de un sistema de producción basado en registros. Sin este registro base es difícil crear estadísticas basadas en registros; con este registro será más fácil continuar y desarrollar otras partes del sistema. Por ende, un proyecto para desarrollar un registro de población estadístico se debe iniciar temprano cuando un país ha decidido desarrollar estadísticas basadas en registros. Una vez creada la primera versión de este registro, el siguiente paso es evaluarlo. Para eso, se pueden utilizar las encuestas por muestreo existentes basadas en marcos de área.

Los registros administrativos de población en los países latinoamericanos visitados tienen dos fallas principales: una cobertura insuficiente de las personas que no han sido registradas y domicilios residenciales obsoletos. Esto se comentó en el **apartado 2.4** y se continuará comentando en el **apartado 5.3**, donde se describe la manera en que el muestreo por área se puede usar para estimar estos errores.



Otro tipo de aseguramiento de calidad es el análisis de usabilidad de cada registro administrativo. Wallgren y Wallgren (2014: 129, 176, 273-277) describen cuatro grupos de indicadores de calidad que se pueden utilizar para este fin.

El primer grupo de indicadores mide la relevancia y se usa para analizar metadatos de cada registro administrativo. Esta información consiste en la documentación creada por la autoridad administrativa responsable de la fuente. Para este análisis deben utilizarse las formas administrativas, instrucciones de soporte y entrevistas con los proveedores de datos. La relevancia de la población, unidades y variables administrativas se evalúan con base en estos metadatos.

El segundo grupo de indicadores mide la precisión y se basa en el análisis de microdatos para cada registro administrativo. Se documenta la calidad de las claves de concordancia, la existencia de duplicados, los valores faltantes y los incorrectos.

El tercero y cuarto grupos de indicadores se basan en comparaciones con otras fuentes y registros del sistema. Los problemas de cobertura e inconsistencia entre variables se analizan con estos indicadores. Laitila, Wallgren y Wallgren (2012) presentan estos cuatro tipos de indicadores y mediante un estudio de caso muestran cómo se pueden utilizar para evaluar la calidad de un registro administrativo.

## 5 Problemas relacionados con el registro de población

La calidad del registro de población estadístico depende de la calidad de los registros administrativos utilizados como entrada. De todas maneras, los registros administrativos de baja calidad pueden compensarse hasta cierto punto con los métodos estadísticos adecuados que permiten combinar la información de otros registros.

Wallgren y Wallgren (2014) analizan las variables que se deben incluir en el registro de población estadístico. También ilustran cómo se actualiza este registro con nueva información. Las variables de identificación, las de referencia con vínculos a otros tipos de unidades estadísticas y las referencias de tiempo<sup>3</sup> con fechas de eventos demográficos son variables importantes dentro del registro de población.

En esta sección, en primer lugar se aborda el sistema nacional de registro, el cual determina cómo se registra a las personas en el registro civil.<sup>4</sup> ¿Cómo se puede mejorar el sistema de registro? Crear un sistema estadístico eficiente basado en registros puede requerir mejorar las rutinas utilizadas para crear y mantener el registro civil.

En segundo lugar se analiza cómo encontrar los duplicados y cómo tratarlos. Una persona puede tener registros en diferentes puntos en el tiempo, y esta información se debe mantener en el sistema del registro de población para poder rastrear los cambios en el tiempo.

Finalmente, se aborda el tema de la cobertura insuficiente en el registro de población y el problema con los domicilios particulares que no se han actualizado. Para monitorear estos problemas de calidad, es necesario llevar a cabo una encuesta por muestreo con base en marcos muestrales basados en área. Para hacer ajustes frente a este tipo de errores pueden usarse métodos similares a los que se utilizan para corregir las estimaciones por falta de respuesta.

3 Se recomienda que las referencias de tiempo tengan el formato AAAAMMDD. Las variables de tiempo definidas de esta forma son de escala ordinaria y pueden utilizarse fácilmente para definir diferentes poblaciones de registro.

4 En todos los países, una autoridad es responsable del registro administrativo de la población. En esta publicación llamamos al registro “registro civil” y a la autoridad “oficina del registro”. Utilizamos el término “número de cédula” para referirse al NIP.

## 5.1 El registro civil y el sistema nacional de estadística

La oficina del registro civil es la autoridad responsable de registrar a las personas y del sistema de números de identidad utilizado en el país. Debido a que la población de personas es la población objetivo más importante, el registro de población estadístico será el más importante en un sistema nacional de estadística moderno. Cuando las estadísticas sociales se basan en registros, puede ser necesario mejorar las rutinas usadas para los registros que maneja la oficina del registro civil. De otra forma, no sería posible crear estadísticas basadas en registros de calidad suficiente. La oficina del registro civil y la ONE deben trabajar juntas en un proyecto a largo plazo para este fin.

En los países nórdicos los registros administrativos de población y los sistemas de números de identidad han sido de buena calidad desde el principio. Cuando estos países empezaron a producir estadísticas basadas en registros en la década de 1960, podían utilizar una sola fuente para el registro de población, ya que su calidad era buena. Además, el sistema de números de identidad estaba bien establecido y fue fácil de usar desde el principio. Para estos países fue posible utilizar métodos sencillos para obtener estadísticas basadas en registros sobre personas.

La situación en muchos países es muy diferente comparada con la de los países nórdicos. Si la calidad del registro civil no es suficiente para fines estadísticos, se deben combinar muchos registros para obtener una cobertura aceptable. Los domicilios particulares en el registro civil pueden tener una baja calidad, lo que significa que para el registro de población estadístico debe utilizarse la información sobre domicilios de otras fuentes. Se pueden usar diferentes sistemas paralelos para identidades y crear duplicados cuando una persona se registra por segunda vez y recibe un nuevo número de identidad.

Los nacimientos, muertes, inmigración, matrimonios y divorcios se registran en el registro civil. En algunos países la emigración, la inmigración y la migración dentro del país no se registran, lo cual da lugar a serios problemas estadísticos. Las estadísticas regionales serán difíciles de obtener si la migración dentro del país no se incluye en el sistema nacional de registro.

### La emigración y la inmigración

El Departamento Administrativo Nacional de Estadística (DANE), que es la ONE de Colombia, ha desarrollado un método para clasificar a las personas como emigrantes o inmigrantes. Se escanean los pasaportes de las personas que salen de Colombia o entran al país. Se las clasifica como personas que

han salido o entrado permanentemente con base en los datos administrativos de un periodo de al menos 12 meses. Al combinar estos datos con los números de identidad, se puede actualizar el registro de población estadístico. Se recomienda que este método también se utilice en otros países.<sup>5</sup> Al final, la mejor alternativa será que cada país actualice su registro civil con información sobre emigración e inmigración.

### Cambio de los hábitos de registro en el país

En Ecuador ha habido conversaciones respecto de cómo se puede mejorar el registro nacional, en las cuales se mencionaron dos alternativas:

- Que la oficina del registro civil sea la responsable de actualizar los domicilios particulares y la información sobre la composición de los hogares.
- Que los municipios sean los responsables de crear y mantener los registros municipales con los domicilios particulares y la composición de los hogares actualizados.

Sin embargo, tomará tiempo implementar los registros que manejan los municipios. Hasta entonces, la oficina del registro civil debe ser la responsable de actualizar los domicilios en el registro civil mientras que el Instituto Ecuatoriano de Seguridad Social será el responsable de actualizar la composición de los hogares.

Se discutió una tercera alternativa durante la visita realizada en 2015. El INEC, la ONE de Ecuador, podría ser el responsable de una plataforma de internet en la cual las instituciones como el Ministerio de Educación, las universidades, el Instituto Ecuatoriano de Seguridad Social, el Instituto Fiscal y otros fueran responsables de actualizar la información sobre la población. Cuando los ciudadanos necesiten el servicio de estas instituciones, ellos mismos deberán proporcionar información actualizada, por ejemplo dónde viven y con quién, a través de la plataforma de internet.

La recomendación de esta publicación se basa en la experiencia de Suecia. En lugar de crear estructuras paralelas, cada una trabajando con datos que son relevantes para el sistema nacional de registro, la oficina del registro civil debería tener toda la responsabilidad. Sin embargo, los municipios y otras autoridades deben cooperar y apoyar a la oficina del registro civil en esta tarea. Para eso, debe ser obligatorio para estas autoridades presentar toda la información

5 Además, Suecia podría utilizar este método, para reducir el exceso de cobertura en su registro de población.

y correcciones solicitadas por la oficina del registro civil en su tarea para mantener y actualizar el registro civil. Para tal fin, en Suecia se presentan alrededor de 1,2 millones de correcciones en la oficina del registro civil cada año.

La ONE debe participar en este trabajo de mejoramiento del sistema de registro y puede contribuir midiendo la calidad y las mejoras en esta. Sin embargo, siempre se debe seguir el principio de tráfico en un solo sentido: los microdatos pueden ir de las autoridades administrativas a la oficina de estadística pero nunca en la dirección opuesta.

### Un registro de los medidores de energía eléctrica como marco muestral

En Colombia y Ecuador se discute si los datos administrativos de las compañías de electricidad sobre los domicilios de facturación de los usuarios pueden utilizarse con fines estadísticos. Los medidores de electricidad están georreferenciados y esta información puede ser de gran ayuda para vincular a las personas con los edificios.

Si la mayoría de los hogares son usuarios de energía eléctrica, entonces los registros de usuarios de las compañías de electricidad se pueden utilizar como un marco muestral de hogares. Este marco muestral se podría usar para generar encuestas por muestreo basadas en registros, lo que podría ser una alternativa rentable al muestreo basado en marcos muestrales de área. El registro de medidores de energía eléctrica también se podría emplear para sustituir la información de vivienda en los censos de población y vivienda por un cuestionario enviado a todos los hogares o a una muestra de hogares.

### Entrega de los datos

El INEC en Ecuador ha recibido la primera entrega de datos del registro civil con el recuento actual de la población al 31 de diciembre de 2014, con actualizaciones trimestrales planeadas. La entrega consistió en dos registros: uno con las personas que se habían registrado y recibido un número de cédula y otro con las personas que tienen un acta de nacimiento pero no han solicitado un número de cédula.

La entrega de datos al INEC es una nueva tarea para la oficina del registro civil y los métodos de entrega se deben mejorar. Solo hubo un registro para cada número de cédula, y se sospecha que solo se entregó la población activa al 31 de diciembre, lo que significa que no es posible crear estadísticas de población retroactivas.

El INEC editó cuidadosamente los datos del registro civil y reportó la estructura de error al personal de la oficina del registro civil. Esta fue la primera experiencia de retroalimentación de calidad por parte de expertos externos.

Cuando el INEC recibió la primera actualización después del primer trimestre de 2015, surgió otro problema: la oficina del registro civil solo pudo enviar la población completa nuevamente, constituida por alrededor de 18 millones de registros, y no solo los registros modificados. Estas experiencias muestran la necesidad de poner toda la atención en esta etapa del sistema nacional de estadística. Entregar datos a la oficina de estadística es una tarea nueva para las autoridades administrativas. Por eso, es importante entender qué tipo de datos se requieren, cómo se deben procesar y seleccionar, y cómo debe la autoridad administrativa usar la nueva retroalimentación de calidad.

## 5.2 Cómo encontrar duplicados

En el INEC los duplicados en el registro de población se consideran un problema importante. Se pueden encontrar duplicados al comparar nombres, fechas de nacimiento o lugares de nacimiento. Por ejemplo, la misma persona aparece registrada dos veces y tiene dos números de cédula o tiene un registro con un número de cédula y otro sin número. El INEC sabía que los registros recibidos de la oficina del registro civil contenían duplicados, pero ¿cómo se pueden encontrar estos duplicados y cómo se deben manejar? La solución se sugiere a continuación.

Los registros de población pueden ser un arduo proceso. El INEC trabaja con registros que constan de 18 millones de entradas. Si se comparan estos 18 millones de entradas con todos los demás registros, es posible involucrar alrededor de 3.240 millones de comparaciones por computadora. Para facilitar el procesamiento de datos, se ha decidido no usar la vinculación de registros, sino que el método propuesto se basa en clasificar los registros y comparar únicamente los consecutivos.

Como se muestra en el **cuadro 3**, el código Nombre contiene los códigos de los nombres; el código Paterno y el código Materno corresponden al primero y segundo apellidos. En el cuadro, los nombres reales se han sustituido por códigos numéricos.<sup>6</sup> De esta manera, María podría recibir el código 1076, González el código 1018 y así sucesivamente. Hay seis variables de identifi-

6 Dado que el ejemplo se basa en datos reales, los nombres que aparecen han sido modificados en esta publicación.

cación con nombres y fechas de nacimiento. En una aplicación real, el lugar de nacimiento también se debería incluir.

**Cuadro 3.** Búsqueda de duplicados al clasificar por identidades

Código Nombre	Código Paterno	Código Materno	Año (AAAA)	Mes (MM)	Día (DD)	Código combinado	Equivale a
...	...	...	...	...	...	...	...
1076	1018	1751	1986	6	10	1076.1018175119860610	0
1076	1451	396	1994	3	9	1076.1451039619940309	0
1077	426	880	1982	4	1	1077.0426088019820401	0
1078	274	1001	1978	3	25	1078.0274100119780325	0
1079	1760	1779	1991	12	3	1079.1760177919911203	0
1079	1854	23	1991	11	15	1079.1854002319911115	0
1079	1854	23	1991	11	15	1079.1854002319911115	1
1080	2198	500	1986	12	5	1080.2198050019861205	0
1081	1631	2337	1988	1	7	1081.1631233719880107	0
...	...	...	...	...	...	...	...

La regla más estricta para duplicados es que si las seis variables son iguales, entonces dos registros se consideran duplicados. Se empieza la búsqueda de duplicados con la regla más estricta:

- Se crea un código combinado al concatenar las seis variables de identidad.
- Todos los registros se clasifican por el código combinado.
- Se comparan los registros subsecuentes: si dos registros subsecuentes tienen códigos combinados iguales, se ha encontrado un duplicado.
- Los dos registros duplicados se quitan para darles un tratamiento especial. En el caso del INEC, estos dos registros tienen números de cédula diferentes; otras variables también pueden diferir, por ejemplo, el estado civil. Entonces el duplicado puede representar información de dos puntos diferentes en el tiempo para la misma persona. Estos dos registros con diferente número de cédula recibirán el mismo número de registro anónimo en la base de datos de identidad.
- Si se acepta un error o diferencia entre los códigos Paterno, Materno, año, mes y día de nacimiento se pueden buscar más duplicados de la misma forma.
- Las pequeñas diferencias en Nombre se pueden manejar computando un índice de similitud. Como ejemplo, los nombres Ester y Esther pueden tener un índice cercano al 100%. Si la similitud está por arriba de cierto límite, por ejemplo el 80%, y las otras variables identificadoras son iguales,

entonces se pueden tratar los registros como duplicados.

- Los duplicados definidos de acuerdo con una regla de similitud especificada se pueden encontrar de esta forma. La información de los dos registros perteneciente al mismo duplicado se puede combinar a través del mismo número de registro anónimo. Los dos registros representan información de la misma persona a partir de diferentes fuentes o puntos en el tiempo.

### 5.3 Cobertura insuficiente y domicilios particulares incorrectos<sup>7</sup>

Los errores de cobertura son un problema que debería recibir mayor atención. Existe una larga tradición de trabajar con errores de muestreo y falta de respuesta. El alza en las tasas de falta de respuesta en las encuestas de la Oficina Nacional de Estadística de Suecia ha recibido recientemente mucha atención en los medios del país. Sin embargo, los problemas de cobertura no han sido discutidos en los medios. El registro de población utilizado por la Oficina Nacional de Estadística de Suecia tiene una sobrecobertura y una cobertura insuficiente de entre el 0,5% y el 1% para ambas.

El registro de empresas utilizado en las estadísticas económicas anuales tenía para el año 2004 una sobrecobertura de alrededor del 8% y una cobertura insuficiente de alrededor del 28% en relación con el número de unidades legales y alrededor del 2% en relación con la rotación total (Wallgren y Wallgren, 2014: 139-140, 222). Estos errores de cobertura también se encuentran en todas las encuestas por muestreo que utilizan estos registros como marcos de muestreo.

Aun cuando los problemas de cobertura se consideren pequeños a nivel general, los errores pueden ser altamente selectivos. Un ejemplo es la cobertura excesiva en el registro de población sueco, que genera serios problemas para las estadísticas por país de nacimiento. El problema aún no ha sido atendido. En un país como Suecia, que cuenta con muchos registros administrativos buenos, estos errores de cobertura se pueden reducir utilizando más fuentes y mejorando los métodos estadísticos. En los países en desarrollo que empezaron recientemente a utilizar registros administrativos, se sugiere utilizar marcos duales para estimar los errores de cobertura; esto es: combinar registros con encuestas por muestreo basadas en marcos muestrales por área. El enfoque de doble marco para encuestas por muestreo fue comentado por

7 Esta sección se basa en Wallgren y Wallgren (2015b).



Ferraz (2015) y Carfagna y Carfagna (2010). También puede utilizarse la técnica de calibración desarrollada por Deville y Särndal (1992) para corregir estimaciones para errores de cobertura.

Los errores de cobertura en los censos con frecuencia se estiman por medio de encuestas después del conteo. El alcance y métodos utilizados para estas encuestas difieren de los que se analizan aquí. En una encuesta posterior al conteo se hacen mediciones repetidas utilizando el mismo marco que en el censo. Los errores de medición encontrados se utilizan para estimar errores por cobertura insuficiente en el censo. Un documento reciente de da Silva, de Freitas y Pessoa (2015) describe la manera en que se usó este método en el censo de 2010 de Brasil.

El **gráfico 6** en el **apartado 2.4** describe las diferentes partes de la población que se incluyen o excluyen del registro. El **gráfico 7** ilustra los tres tipos de encuestas que se discuten aquí: por registros, por muestreo basada en registros, por muestreo basada en un marco muestral por área. Ferraz (2015) considera el caso en que una muestra basada en listas o en registros<sup>8</sup> y una muestra basada en un marco muestral por área que contiene las mismas variables se combinan durante la etapa de estimación. Todas las variables se recopilan para ambas muestras.

Aquí se considera un caso diferente que es importante para los países que han empezado a utilizar registros administrativos para crear registros estadísticos. Es posible que los errores de cobertura y los problemas de los domicilios particulares que no han sido actualizados hagan necesario combinar un registro estadístico con una muestra basada en área. Las estadísticas obtenidas se basan principalmente en el registro o en encuestas por muestreo basadas en marcos creados con el registro. Es deseable producir estimaciones por municipio con el registro. La muestra basada en área se utiliza para estimar los errores de cobertura a nivel nacional y regional y para diferentes categorías de municipios, personas u hogares. También es posible ajustar o corregir por errores de cobertura con condiciones de calibración basadas en algunas variables del registro.

Supóngase que se quiere producir estadísticas regionales con el registro de población creado por una ONE. El **cuadro 4** enumera algunas de las variables del registro. Los municipios se han clasificado en diferentes categorías con base en el grado de urbanización y quizá en el tamaño relativo de la población indígena. Se considera que estas variables pueden influir en el tamaño

8 En este artículo se prefiere el término registro en lugar de lista.

de los errores de cobertura para un municipio. Tal vez sea posible clasificar los hogares o las personas en distintas categorías con diferente propensión a moverse a otro domicilio particular. La edad, nivel de educación y actividad económica son variables que se pueden utilizar aquí.

**Cuadro 4.** Variables en las tres encuestas diferentes

Variables en el registro	Variables en la muestra basada en marco de área	Variables en la muestra basada en registro
Número de identidad, <i>Idnr</i>	Si la unidad no está en el registro: <i>Idnr</i> = faltante	Número de identidad, <i>Idnr</i>
Domicilio particular	Otras variables: igual que en el registro	Otras variables: igual que en el registro y variables estadísticas para la encuesta
A partir de la fecha de residencia	Es posible agregar variables estadísticas	
Distrito		
Municipio		
Categoría de municipio		
Sexo		
Edad		
Nivel de educación		
Actividad económica		
Categoría de persona u hogar		

Para una muestra de área, los entrevistadores recaban las mismas variables del registro. Puede haber personas en la muestra que no se encuentran en el registro o que tienen diferentes domicilios en el registro. También es posible encontrar una cobertura insuficiente o excesiva en diferentes niveles regionales: municipio incorrecto, distrito incorrecto y personas no incluidas en el registro.

### Estimación de errores de cobertura

Se pueden calcular varias medidas de calidad con base en comparaciones entre la muestra de área y el registro: los porcentajes de personas con domicilio particular incorrecto, el retraso para reportar un domicilio nuevo y las tasas de cobertura insuficiente a nivel regional y nacional. El trabajo a largo plazo en las mejoras del sistema nacional de registro se debe monitorear con estas medidas.

El **cuadro 5** ilustra los diferentes tipos de errores de cobertura. A nivel nacional, la muestra de área proporciona una cobertura insuficiente estimada de 72.000 personas en el registro. La estimación de la muestra basada en área

es que 80.000 personas viven realmente en el distrito A, pero están registradas en el distrito B. Estas 80.000 personas son una cobertura insuficiente estimada de la subpoblación que vive en el distrito A y al mismo tiempo una sobrecobertura de la subpoblación que vive en el distrito B. Las personas que han salido del país corresponden a otra categoría de sobrecobertura que no se puede detectar con los métodos analizados aquí.

**Cuadro 5.** Ejemplos de errores de cobertura, estimaciones por muestra utilizando marcos muestrales de área de la población por distrito

Distrito según los datos del registro							
	Distrito	A	B	C	No está en el registro	Todos	Subcobertura
Distrito según la muestra por área	A	2.870.000	80.000	110.000	19.000	<b>3.079.000</b>	209.000
	B	16.000	3.110.000	40.000	27.000	<b>3.193.000</b>	83.000
	C	2.000	10.000	1.530.000	26.000	<b>1.568.000</b>	38.000
	<b>Todos</b>	<b>2.888.000</b>	<b>3.200.000</b>	<b>1.680.000</b>	72.000	<b>7.840.000</b>	330.000
Sobrecobertura		18.000	90.000	150.000			

**Cuadro 6.** Estimaciones no corregidas y corregidas

Distrito	Estimados originales	Estimados corregidos
A	2.888.000	3.079.000
B	3.200.000	3.193.000
C	1.680.000	1.568.000
<b>Todos</b>	<b>7.768.000</b>	<b>7.840.000</b>

En los **cuadros 5 y 6**, los estimados de registro originales de la población en los distritos se pueden comparar con los estimados de la muestra por área. En el registro el distrito A está subestimado y los distritos B y C, sobreestimados. Se pueden comparar los estimados de registro y los de muestra por área para otras categorías en forma similar.

### Estimado de doble marco

Se considera que la muestra por área puede proporcionar estimaciones confiables a nivel regional y que los municipios definen el nivel regional deseado para las estimaciones basadas en el registro. La estimación con el enfoque de doble marco tiene algunas deficiencias, que harán que los estimadores de

doble marco analizados en Ferraz (2015) sean difíciles de utilizar para el caso que se analiza aquí:

- No se considera posible que la muestra por área se pueda diseñar de manera tal que se puedan obtener los estimados de doble marco para los municipios. Los límites de costos harían esto imposible.
- Existirán inconsistencias entre los datos del registro y las medidas tomadas por los entrevistadores. Por ejemplo, algunas personas dirán que han terminado la escuela secundaria pero según los datos del registro de escuelas, no lo han hecho.
- Los datos del registro regional y de las encuestas por muestreo basadas en registros estarán influenciadas por los errores de sobrecobertura. Los estimadores de doble marco solo corregirán la subcobertura. Se detectará la sobrecobertura del registro en la muestra basada en marcos de área cuando los entrevistadores se reúnan con personas que han cambiado de domicilio en la muestra. En esos casos, los entrevistadores comprenderán que estos domicilios son incorrectos como domicilios particulares de estas personas pero será difícil determinar si el distrito o el municipio es correcto o no.

La conclusión es que los métodos de doble marco discutidos por Ferraz (2015) no son apropiados cuando se tiene tanto subcobertura como sobrecobertura para campos importantes de estudio. Los errores en el registro (marco de listas) en relación con las variables de clasificación utilizadas para definir los campos de estudio, generarán este tipo de errores de cobertura. La ubicación de la residencia de personas y establecimientos son ejemplos de dichas variables de clasificación. Por lo anterior, es necesario un enfoque diferente de doble marco si se quieren generar estadísticas basadas en el registro de población y corregir los errores de cobertura.

### Ajuste por errores de cobertura a través de la calibración de pesos

El enfoque alternativo de doble marco propuesto en este apartado utiliza procedimientos similares al método de calibración usado para corregir la falta de respuesta en encuestas por muestreo. La muestra por área analizada en el apartado anterior se puede utilizar de dos formas:

- Para proporcionar estimados de los errores de cobertura y en las variables de clasificación básica en el registro, de acuerdo a lo descrito en el apartado anterior sobre el estimado de doble marco.
- Para proporcionar estimaciones que puedan utilizarse como condiciones de calibración para ajustar o corregir errores de cobertura.

El **cuadro 7** muestra la estructura del registro de población. Las variables de las columnas 2 a 6 se pueden utilizar para definir las condiciones de calibración si se quiere ajustar por errores de cobertura. Si se elige utilizar únicamente el distrito y la información sobre el número de personas de cada distrito como condiciones de calibración, los pesos originales  $d_i$  de la columna 7 se transformarán en los pesos calibrados  $w_i$  de la columna 11.

**Cuadro 7.** Siete registros del registro de población

Número de identidad	Sexo	Edad	Municipio	Categoría	Distrito	$d_i$	$x_1$	$x_2$	$x_3$	$w_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
100002	2	99	1130	2	B	1	0	1	0	0,998
100003	2	33	1224	4	C	1	0	0	1	0,933
100012	1	73	1240	2	C	1	0	0	1	0,933
100013	1	19	1134	3	B	1	0	1	0	0,998
100019	1	64	1243	4	C	1	0	0	1	0,933
100023	1	24	1040	2	A	1	1	0	0	1,066
100026	1	72	1025	1	A	1	1	0	0	1,066
...	...	...	...	...	...	...	...	...	...	...

Si se suma la columna 7 por distrito, se obtendrán los estimados originales sin corregir del **cuadro 6**. Si se suma la columna 11 por distrito, se obtendrán los estimados corregidos del **cuadro 6**.

En este caso es fácil calcular los pesos calibrados. Los pesos  $w_i$  para los registros señalados como distrito A en el registro serán  $3.079/2.888 = 1,066$  y los pesos para los otros distritos se calculan de manera similar. A partir de las columnas 8 a 10, se pueden derivar los mismos pesos con álgebra matricial. Se deben analizar los datos de la muestra basada en marcos de área y se pueden utilizar las diferencias relacionadas con los errores de cobertura para diferentes categorías de sexo, edad, categoría de municipio y distrito para construir un mejor modelo que pueda utilizarse para ajustar o corregir los estimados por errores de cobertura. Este trabajo metodológico será muy similar al que se realiza cuando se ajustan los datos por falta de respuesta en las encuestas por muestreo.

Si existen más registros estadísticos con datos sobre, por ejemplo, educación, empleo o ingresos, estos deben utilizar la población definida por el registro de población como sus poblaciones de registro. Los pesos calibrados  $w_i$  también deben utilizarse para la estimación de estos registros. Estos regis-

tros también serán corregidos para errores de cobertura de esta forma. Estas correcciones son consistentes y las estadísticas generadas con todos estos registros serán coherentes.

### Encuestas por muestreo basadas en registros

Las encuestas por muestreo que utilizan el registro de población como marco muestral también deben ser corregidas por errores de cobertura en el registro de población. Al ajustar los pesos utilizados para producir las estimaciones en la encuesta por muestreo, las estimaciones se corrigen tanto por falta de respuesta como por errores de cobertura si se utilizan los totales de los registros que han sido corregidos por errores de cobertura. De esta forma, se combina la información de la muestra por área con la información de la muestra basada en registros a fin de obtener estimaciones corregidas o ajustadas por errores de cobertura. Se ha vinculado un marco de área con un marco de lista en los términos utilizados en Ferraz (2015) para producir las mejores estimaciones posibles. Sin embargo, el problema estadístico difiere del problema estudiado por Ferraz, ya que aquí se consideran las estadísticas basadas en registros y se utiliza la muestra por área para corregir las estimaciones basadas en registros.

### Discusión

Muchos países tienen registros de población, registros de empresas y registros de explotación agrícola que tienen problemas tanto de subcobertura como de sobrecobertura. La información básica de registros, como domicilio particular o actividad económica, puede estar desactualizada o ser incorrecta. Estos factores de calidad deben monitorearse por muestreo utilizando marcos por área. La calidad de las estimaciones basadas en registros también se puede mejorar calibrando los ponderadores. Las variables de los registros deben utilizarse para definir las condiciones de calibración que reducirán los errores de cobertura. Al final, los sistemas nacionales de registro también se deben mejorar para reducir gradualmente los problemas de cobertura.

En la actualidad se publican estimaciones de encuestas por muestreo con tasas de no respuesta del 40% o incluso mayores. Puede hacerse porque hay confianza en los métodos empleados para ajustar estimaciones por no respuesta con la información de los registros que se usan como marcos. ¿Podrían también ajustarse las estimaciones basadas en registros cuando se tiene un 40% de subcobertura? Esta idea resulta nueva y sorprendente dado que en Suecia la subcobertura es cuando mucho del 0,5% al 2%, la cual se considera tan pequeña que se ignora. En primer lugar, es necesario aclarar que

ahora se sabe que los errores de subcobertura pueden ser muy selectivos y que una sobrecobertura o subcobertura de entre el 0,5% y el 2% no se debe despreciar ya que puede generar errores graves. En segundo lugar, es cierto que cuando estudiamos estadística tiempo atrás, se enseñaba que si la tasa de falta de respuesta era mayor al 5%, la encuesta por muestreo era inútil!

La conclusión a la que se puede arribar es que las actitudes hacia los errores en las encuestas cambian con el tiempo. Si se tiene confianza en los métodos para ajustar las estimaciones por no respuesta o por errores de cobertura, es posible hacer correcciones y publicar. Corregir las estimaciones por no respuesta y por errores de cobertura con el método que aquí se propone requiere de métodos muy similares. Por lo tanto, se recomienda lo siguiente:

- combinar los datos de registro con datos de una muestra basada en marco de área,
- analizar los problemas de cobertura y definir las condiciones de calibración adecuadas, y
- corregir las estimaciones y publicar si se tiene confianza en el método de corrección.

## 6 Conclusiones

Aquí se retoman las conclusiones del **apartado 2.5** y se agregan algunos comentarios: si se usan registros administrativos para la producción de estadísticas, se pueden ahorrar recursos y mejorar la calidad. Los censos pueden remplazarse por encuestas por registros, y se pueden desarrollar nuevos tipos de estadísticas regionales y longitudinales. El sistema nacional se puede cambiar a un sistema estadístico integrado con registros estadísticos consistentes y coherentes y encuestas por muestreo. Durante el periodo de transición quizá sea necesario combinar estadísticas basadas en registros con estimaciones de muestras por área.

Si esta visión se convierte en realidad, se requerirá una cooperación permanente entre ministerios, ONE y autoridades. Tanto la alta dirección como el personal que trabaja con los registros administrativos deben participar en esta cooperación. Además, el trabajo dentro de la ONE debe organizarse de una nueva manera. El modelo de proceso de producción presentado en el **apartado 4.1** muestra cómo se puede organizar el trabajo en equipo con diferentes unidades de registro. Se deben definir responsabilidades claras para reducir la duplicación del esfuerzo y mejorar la consistencia. Los administradores de la ONE deben participar en esta coordinación.

Los datos se pueden usar de manera eficiente en un sistema integrado de registros, lo cual disminuye los costos y permite a su vez la comparación de las poblaciones y las variables, lo que hace posible encontrar y reducir errores. Sin embargo, este principio que parece sencillo es muy difícil de implementar. Las personas que trabajan con diferentes encuestas tienen distintas áreas de interés y competencias. Además, más de un administrador estará involucrado; cada uno con su propio presupuesto. La encuesta mensual de negocios por muestreo y el registro anual con datos de empresas en la Oficina Nacional de Estadística de Suecia pueden ilustrar esto. En noviembre del año  $t$  se define la población marco para la encuesta mensual de negocios por muestreo para enero del año  $t+1$ . Las pequeñas empresas no se incluyen en este marco. Se llevan a cabo encuestas por muestreo similares de febrero a diciembre del año  $t+1$ .

Durante septiembre del año  $t+2$ , una unidad de registro de la Oficina Nacional de Estadística de Suecia recibe declaraciones de ingresos para el año  $t+1$  de todas las empresas. Entonces se pueden comparar las poblaciones y variables de las 12 encuestas mensuales por muestreo con el registro empresarial anual. Los errores de marco en las encuestas por muestreo pueden ser graves, por



lo que sería importante monitorear estos errores y tratar de encontrar métodos para reducirlos. Sin embargo, el personal de la unidad de registro no cree que pueda utilizar los datos de la encuesta por muestreo, y el personal de la unidad de encuestas por muestreo no cree que pueda utilizar los datos del registro, así que nadie analiza la calidad mediante la comparación entre estos dos conjuntos de datos. Como resultado, los errores pasan desapercibidos y se mantienen año tras año.<sup>9</sup> El paradigma de una encuesta a la vez prevalece y el sistema de registro no se utiliza en su máximo potencial.

Estas experiencias explican por qué se recomienda organizar una unidad de registro central. Esta unidad debe supervisar el sistema de producción y medir y reducir los errores e inconsistencias. Las oficinas líderes de estadística en países como Canadá y Estados Unidos aún tienen sistemas tradicionales de estadística, en los cuales el paradigma de una encuesta a la vez es relevante. Estos institutos influyen en el pensamiento estadístico de otros países, lo cual tal vez explica por qué los metodólogos y administradores en muchos países aún no han entendido el enfoque de sistemas.

9 Wallgren y Wallgren (2014: 130-140) analizan los errores de cobertura que se pueden detectar y corregir.

# Referencias

- Carfagna, E. y A. Carfagna. 2010. Alternative Sampling Frames and Administrative Data: Which Is the Best Data Source for Agricultural Statistics? En R. Benedetti, M. Bee, G. Espa y F. Piersimoni (eds.), *Agricultural Survey Methods*. Chichester, Reino Unido: John Wiley & Sons, Ltd.
- CEPAL (Comisión Económica para América Latina y el Caribe). 2020. Ley Genérica sobre Estadísticas Oficiales para América Latina. Santiago de Chile: CEPAL. Disponible en: [https://repositorio.cepal.org/bitstream/handle/11362/45253/1/S2000046\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/45253/1/S2000046_es.pdf).
- da Silva, A. D., M. de Freitas y D. Pessoa. 2015. Assessing Coverage of the 2010 Brazilian Census. *Statistical Journal of the IAOS* 31: 215-25.
- Deville, J. y C. E. Särndal. 1992. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association* 87: 376-82.
- FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). 2010. Global Strategy to Improve Agricultural and Rural Statistics. Informe No. 56719-GLB. Washington, D.C.: Banco Mundial.
- Ferraz, C. 2015. Linking Area and List Frames in Agricultural Surveys. Documento preparado para la FAO. Disponible en: <http://www.fao.org/3/ca6438en/ca6438en.pdf>.
- Laitila, T., A. Wallgren y B. Wallgren. 2012. Quality Assessment of Administrative Data: Data Source Quality. Parte 2 del tercer producto final del paquete de trabajo 4 de BLUE Enterprise and Trade Statistics Project. Disponible de los autores en: [ba.statistik@telia.com](mailto:ba.statistik@telia.com).
- Oficina de Estadísticas de Canadá. 2009. Statistics Canada Quality Guidelines (5ta edición).
- Stender P., T. Thorsen y H. Andersen. 2015. Microdata Integration for Labour Market Account. *Statistical Journal of the IAOS* 31: 249-256.
- Wallgren A. y B. Wallgren. 2012. Estadísticas basadas en registros. Aprovechamiento estadístico de datos administrativos. Ciudad de México, México: INEGI.

- . 2014. *Register-Based Statistics: Statistical Methods for Administrative Data* (2da edición). Chichester, Reino Unido: John Wiley & Sons, Ltd.
- . 2015a. Quality Assessment in Production Systems with Registers and Sample Surveys. *Statistical Journal of the IAOS* 31: 241-47.
- . 2015b. Frames and Populations in a Register-Based Statistical System. Documento presentado en ISI 2015, Rio de Janeiro, 26 al 31 de julio.
- . 2016. Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos. Aguascalientes, México: INEGI. Disponible en: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825086206>.
- . 2021 (de próxima publicación). *Register-Based Statistics: Registers and the National Statistical System* (3ª edición). Chichester, Reino Unido: John Wiley & Sons, Ltd.



