

# Hacia un sistema estadístico integrado y basado en registros

Anders Wallgren & Britt Wallgren  
BA Statistiksystem AB, ex miembros de Statistics Sweden  
correo electrónico: [ba.statistik@telia.com](mailto:ba.statistik@telia.com)

## Introducción

Muchos países ya han comenzado a utilizar sus registros administrativos con propósitos estadísticos. En esta transición, las oficinas nacionales de estadística, principales productores de estadísticas oficiales de un país, deben ajustarse y cambiar sus métodos de sondeo y recolección de datos. Los llamados registros estadísticos en los países que hicieron la transición se han convertido en los marcos muestrales de las encuestas por muestreo y las principales fuentes de variables estadísticas para diseñar los censos. La transición es un proceso gradual en el cual el sistema estadístico se asienta completamente en el uso intensivo de los registros administrativos disponibles en el país. Los países nórdicos ya hicieron la transición de un sistema estadístico basado en marcos muestrales de área a un sistema basado completamente en registros.

En un sistema basado en registros, todas las *encuestas por muestreo* se basan en marcos creados por registros estadísticos. Asimismo, en el caso de los *censos*, como el Censo de Explotación Agrícola, el Registro de Explotación de Suelo se convierte en el marco sobre el cual se diseña el muestreo y se envían todos los cuestionarios a las granjas seleccionadas para obtener la información no disponible en el registro. Este mecanismo describe un nuevo tipo de encuesta, aquella que se basa completamente en los microdatos del sistema de registros estadísticos. Lo mismo ocurre con el tradicional del Censo de Población y Vivienda, el cual viene a ser reemplazado por algunas de estas *encuestas basadas en registros*. Los registros estadísticos también se pueden utilizar para crear un *sistema integrado de registros*. En dicho sistema los datos interinstitucionales se mueven de manera más eficiente, y la consistencia y coherencia se convierten en los principales parámetros de la calidad del dato estadístico, todo lo cual puede mejorar sustancialmente aprovechando el nuevo enfoque sistémico.

En este trabajo exponemos *por qué* los países deberían utilizar los registros administrativos para fines estadísticos y *que* obstáculos deben sortearse para su implementación. También explicamos *cómo* se deberían utilizar dichos registros, *cómo* debería realizarse la transición de un sistema estadístico basado en un marco de área a un sistema basado en registros, y *qué* aspectos estadísticos y metodológicos se deben tener principalmente en cuenta.

Este artículo está basado en nuestros libros sobre estadística basada en registros y en algunos artículos de conferencias que presentamos en NTTS 2011, JSM 2011, ISI 2013 e ISI 2015. De igual forma, está basado en las experiencias que tuvimos a partir de nuestras visitas a algunos países de Latinoamérica entre 2011 y 2015: Colombia, Ecuador, México y Perú. Las conversaciones que hemos sostenido con nuestros colegas en estos países han sido muy importantes para este trabajo.

## Contenido

<b>1</b>	<b>Conceptos básicos.....</b>	<b>3</b>
1.1	¿Qué es una encuesta estadística? .....	3
1.2	¿Qué es un registro? .....	3
1.3	¿Qué es una encuesta por registro? .....	4
<b>2</b>	<b>¿Cuándo y por qué deberíamos usar registros administrativos? .....</b>	<b>7</b>
2.1	El sistema nacional de estadística – dos modelos diferentes.....	7
2.2	¿A qué nos referimos con un sistema integrado? .....	7
2.3	Comparación de costos y calidad .....	11
2.4	Métodos combinados – muestreo por zona y por registros .....	12
2.5	Conclusiones: ¿cuándo y por qué deberíamos usar registros administrativos? .....	13
<b>3</b>	<b>¿Cómo deberíamos usar los registros administrativos? .....</b>	<b>14</b>
3.1	Creación de números de identidad para proteger la privacidad.....	14
3.2	¿Cómo se pueden utilizar los registros administrativos?.....	17
3.3	Cuatro maneras de utilizar un registro administrativo.....	18
3.4	Modelo del proceso de producción y modelo del sistema de registros .....	19
<b>4</b>	<b>¿Cómo hacer la transición del sistema antiguo al nuevo?.....</b>	<b>21</b>
4.1	Modelo de proceso de producción.....	21
4.2	El modelo de sistema de registro.....	26
4.3	Aseguramiento de la Calidad .....	29
<b>5</b>	<b>Problemas relacionados con el Registro de Población.....</b>	<b>30</b>
5.1	El Registro Civil y el sistema nacional de estadística .....	30
5.2	Cómo encontrar duplicados.....	33
5.3	Cobertura insuficiente y domicilios residenciales incorrectos .....	34
<b>6</b>	<b>Conclusiones .....</b>	<b>40</b>
	<b>Referencias .....</b>	<b>41</b>

# 1 Conceptos básicos<sup>1</sup>

Dos principios constituyen la base de nuestros libros sobre estadística basada en registros (Wallgren y Wallgren 2012 y 2014): el *enfoque de encuesta* para datos administrativos y el *enfoque de sistemas*. El enfoque de encuesta se refiere a que tratamos estimaciones, estimadores y calidad como en un libro sobre encuestas por muestreo; mientras que el enfoque de sistema se construye alrededor del concepto de sistema de registros. Igualmente, tratamos el sistema de producción dentro de una oficina nacional de estadística y el papel de los registros administrativos en el diseño y desarrollo de dicho sistema. En esta sección presentamos los siguientes conceptos básicos:

- ¿Qué es una encuesta estadística?
- ¿Qué es un registro?
- ¿Qué es una encuesta por registro y cuál es la diferencia entre registros administrativos y registros estadísticos?

## 1.1 ¿Qué es una encuesta estadística?

Éste es un término esencial utilizado por los estadísticos en todas las oficinas nacionales de estadística. Sin embargo, para muchos estadísticos el término *encuesta* es sinónimo de *encuesta por muestreo*. Esto puede causar confusión cuando hablamos de encuesta basada en registros administrativos. Para evitar dicha confusión, nos basaremos en la distinción que usa Statistics Canada (2009) de tres tipos de encuestas.

El término encuesta es usado genéricamente para abarcar cualquier actividad en la que se recopila o se obtienen datos estadísticos.

De acuerdo con los lineamientos de calidad de Statistics Canada, esta incluyen:

- |   |  |
|---|--|
| <b>Censo</b>                                      | 1. Un <i>censo</i> , el cual busca recopilar datos de todos los miembros de una población;   |
| <b>Encuesta por muestreo</b>                      | 2. Una <i>encuesta por muestreo</i> , en la cual los datos son recopilados de una muestra (usualmente aleatoria) de los miembros de la población;          |
| <b>Encuesta por registro</b><br>(nuestro término) | 3. Recolección de datos de <i>registros administrativos</i> , en el cual los datos provienen de registros originalmente creados con fines no estadísticos. |

Mientras una encuesta basada en muestreo recopila datos de una muestra cuyas unidades fueron seleccionados al azar con el fin de construir estimadores de la población, una encuesta por registro recopila los registros de varias fuentes administrativas para generar los llamados registros estadísticos, sobre los cuales se calculan los parámetros de la población.

## 1.2 ¿Qué es un registro?

Un registro administrativo corresponde a la información recolectada por una autoridad administrativa, el cual es almacenado, procesado y analizado por la propia entidad que lo recolecta. La recolección de los registros se puede hacer para todas las unidades de análisis (personas, empresas, viviendas), estas unidades, también llamadas objetos, usualmente tienen un identificador, el cual permite administrar el registro. La siguiente definición es válida para registros estadísticos y administrativos:

<sup>1</sup> La Sección 1 de este artículo se basa en la Sección 1.5 de Wallgren y Wallgren (2014).

Un *registro* tiene como objetivo ser una lista completa de los objetos de un grupo específico de la población; no obstante, es frecuente que falten datos de algunos objetos debido a deficiencias en la calidad (cobertura e inconsistencia).

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 de cada objeto.

– *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.

Aquí usaremos el término *registro*.

Los siguientes son ejemplos de registros:

- El registro civil o nacional de población produce registros de ciudadanía, nacimientos y muertes.
- La autoevaluación 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 debe entregar un informe mensual a la autoridad tributaria. Esto genera registros mensuales de todas las empresas que presentan dicho informe. En el caso de las empresas más pequeñas, se pueden obtener 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 el archivo contiene las identidades de las personas.

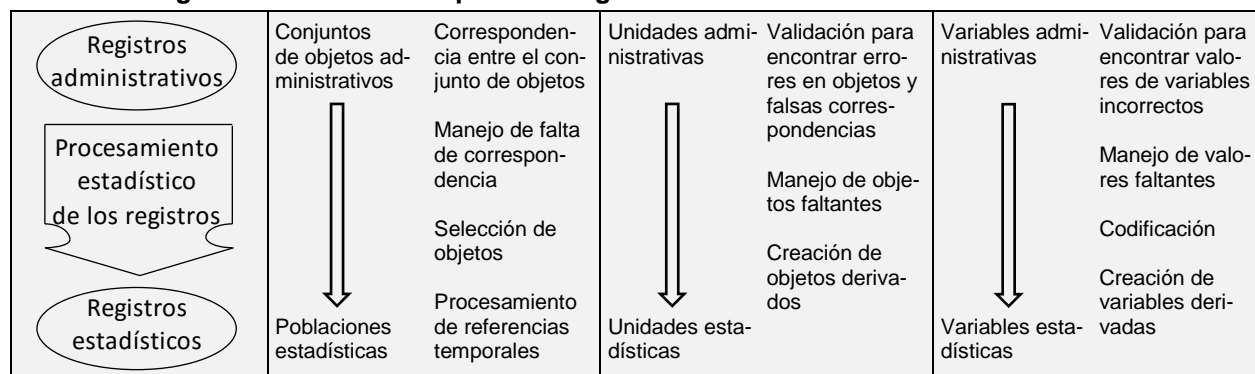
### 1.3 ¿Qué es una encuesta por registro?

En la Sección 1.1 establecimos que en una encuesta por registro se crean uno o más registros estadísticos usando registros o archivos administrativos. Los datos administrativos originales se generan en sistemas públicos administrativos. Las definiciones de los grupos de objetos, los objetos o las unidades administrativas, así como las variables, se adaptan a los propósitos administrativos. Cada autoridad lleva a cabo procesos de control, corrección y de otro tipo, de acuerdo con sus objetivos administrativos.

En general no se recomienda generar estadísticas directamente de los registros administrativos recibidos, ya que éstos no cumplen algunos requisitos estadísticos. Los grupos y definiciones de los objetos y sus variables requieren un trabajo previo de validación. Como regla, siempre es necesario realizar una serie de procedimientos que permitan transformar los registros administrativos en unidades estandarizadas (registros estadísticos) que cumplan con los requisitos estadísticos para medir la población, sus objetos y sus variables.

El Gráfico 1 muestra tres componentes importantes de este trabajo. Hemos encontrado que las personas tienden a usar los conceptos administrativos tal como están; en algunos casos esto puede ser aceptable, pero en otros no. Los tres aspectos sobre cómo definir la población, las unidades y las variables de un registro estadístico son importantes para la calidad de las estadísticas que se producirán a partir del registro estadístico recién creado. El proceso de registros, el cual 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. Registros estadísticos a partir de registros administrativos**



### ***Grupos de objetos administrativos – poblaciones estadísticas***

Puede crearse un Registro de Población estadístico combinando cuatro fuentes: Registro Civil de personas con número de identidad, Registro Civil de personas sin números de identidad, Registro de Nacimientos, Registro de Defunciones y Registro de Migración. La población estadística en un punto específico de tiempo puede crearse usando referencias temporales, como fechas de nacimiento, muertes, inmigración y emigración. Observe que el Registro Civil administrativo y el Registro de Población estadístico creado por la oficina de estadística serán diferentes. El inventario de personas será diferente y las variables en los dos registros también. Por ejemplo, si el domicilio en el registro estadístico se basa en otras fuentes además del domicilio en el 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 de un sistema de seguridad social para crear tres registros estadísticos: trabajo, empresas y personas. Todos los empleadores en el sistema registran las transacciones relacionadas con todos los empleados, que realizan cada mes.

En el Gráfico 2, se usan dos tipos de número de identidad: números de cédula que son números de identidad personal (PIN) y números RUC que son números de identidad laboral (BIN).

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 al instituto estadístico, se crearon los registros estadísticos 2, 3a y 3b. Las unidades administrativas *Transacciones* en el registro 1 se transforman en las unidades estadísticas *Trabajos* en el registro 2, *Empresas* en el registro 3a y *Personas* en el registro 3b.

**Gráfico 2. Unidades estadísticas a partir de unidades administrativas**

3b.Registro de población			2.Registro de trabajo – empleos				1. Transacciones mensuales			
Número de cédula	Días trabajados	Pago de sueldo	Número de cédula	Número RUC	Días trabajados	Pago de sueldo	Número de cédula	Número RUC	Días trabajados	Pago de sueldo
11111111	30	2100	11111111	111	30	2100	11111111	111	10	700
22222222	30	2800	22222222	111	10	1000	11111111	111	5	350
			22222222	222	20	1800	11111111	111	-10	-700
							11111111	111	-5	-350
							11111111	111	30	2100
							22222222	111	10	1000
							22222222	222	10	900
							22222222	222	10	900

3a. Registro empresarial			
Número RUC	Días trabajados	Pago de sueldo	Salario por día
111	40	3100	77.5
222	20	1800	90.0

### **Variables administrativas – variables estadísticas**

Hay dos variables administrativas en el Gráfico 2, *Días trabajados por transacción* y *Pago de sueldo por transacción* en el registro administrativo 1 con transacciones mensuales. En los registros estadísticos 2, 3a y 3b hemos derivado las variables estadísticas *Días trabajados* y *Pagos de sueldo por trabajo, empresa y persona*. El registro 3a también ha derivado la variable estadística *Sueldo por día por empresas*.

En una encuesta por muestreo, definimos las unidades estadísticas, hogares o establecimientos. Los entrevistadores revisan que el tipo de unidades que se pretenden estudiar sean realmente las entrevistadas. Las variables estadísticas están definidas por las preguntas y las definiciones en el cuestionario. En una encuesta por registro, debemos usar las unidades administrativas y las variables administrativas. A veces podemos agregar unidades administrativas y crear unidades estadísticas, como en el Gráfico 2. Podemos usar las variables administrativas para crear variables estadísticas derivadas, como las variables en la matriz de datos 2, 3a y 3b del Gráfico 2.

## 2 ¿Cuándo y por qué deberíamos usar registros administrativos?

Cuando una oficina nacional de estadística 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 continúan; no obstante, en este nuevo modelo los registros administrativos también se utilizan con propósitos estadísticos, y se sumarán nuevas oportunidades para el sistema de producción. En esta sección analizamos algunas consecuencias para el sistema de producción en el caso que los registros administrativos estén disponibles para la producción estadística

### 2.1 El sistema nacional de estadística – dos modelos diferentes

En esta sección comparamos dos tipos diferentes de sistemas nacionales de producción: el *sistema tradicional* en el que no se utilizan registros administrativos para la producción estadística y el *sistema basado en registros*, en el cual muchos registros estadísticos se crean con base en los registros de varias fuentes administrativas. Actualmente, muchos países están en alguna parte entre estos dos extremos, de manera que los registros administrativos se utilizan en cierta medida; no obstante, el sistema nacional de producción todavía no se basa completamente en registros.

**Gráfico 3. Comparación entre el sistema de producción tradicional y el basado en registros**

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

El Gráfico 3 muestra que las principales diferencias metodológicas entre los dos sistemas de producción son: *i)* los tipos de marcos, y *ii)* los métodos de recolección de datos. Los países sin un registro estadístico de población pueden clasificarse como países que tienen un sistema tradicional; mientras que los países que llevan a cabo censos de población y vivienda basados completamente en registros pueden clasificarse como 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 registro se puede basar en la existencia de un registro estadístico de población. Si la oficina nacional de estadística ha sido capaz de crear un registro de población, será relativamente fácil crear los otros registros.

### 2.2 ¿A qué nos referimos con un sistema integrado?

La organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) ha lanzado una Estrategia Global para el Mejoramiento de las Estadísticas Agropecuarias y Rurales en los países en desarrollo. FAO (2010) trata el tema de los sistemas *integrados* de estadística a nivel nacional. Seleccionamos la siguiente cita de dicho informe y marcamos el texto que está íntimamente relacionado con aquello a lo que nos referimos como sistema integrado.

## ¿Cuándo y por qué deberíamos usar registros administrativos?

En la integración y la metodología...se consideran las dimensiones de calidad que incluyen la relevancia, la exhaustividad, la exactitud, la oportunidad, la accesibilidad, la coherencia, y la equivalencia.

El proceso para mejorar las estadísticas agropecuarias debe empezar 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, exactos, coherentes y comparables.

Uno de los defectos de los sistemas estadísticos actuales, tanto en países industrializados como en los 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.

A menudo, las encuestas se realizan sobre bases ad hoc sin conexiones con un marco maestro de muestreo o con el uso de unidades georreferenciadas para la recopilación de datos. Por lo tanto, resulta difícil (si no imposible) integrar esos datos, provenientes de encuestas diferentes, para analizarlos en profundidad mediante tabulaciones cruzadas de las variables.

Los sistemas integrados de estadística pueden resolver muchos de estos problemas, al evitar la duplicidad de esfuerzos, prevenir la comunicación de estadísticas opuestas, y asegurar el mejor uso de los recursos. Los conceptos, definiciones y clasificaciones se normalizan, al admitir la recopilación más sistemática de datos a través de fuentes.

Las partes de un sistema integrado encajan perfectamente entre ellas. En el caso de un sistema estadístico, esto significa que las poblaciones y las variables de distintas encuestas encajan entre ellas. La exhaustividad, coherencia y equivalencia son factores de calidad importantes. Los organismos gubernamentales e institutos que llevan a cabo encuestas estadísticas deberán también coordinar su recopilación de datos para producir estadísticas oficiales. FAO también requiere que los microdatos obtenidos de diferentes encuestas puedan combinarse o integrarse a un nivel macro, mediante la vinculación de registros, con el objetivo de lograr tabulaciones cruzadas de variables. Según FAO, el presente sistema tiene varias dificultades. La falta de cooperación entre los distintos actores lleva a conflictos en la publicación de estadísticas y a la duplicación de esfuerzos, lo que significa que no se están usando de manera eficiente los recursos económicos y humanos disponibles.

En Wallgren y Wallgren (2014), tratamos el concepto de *sistema de registros*. En el Capítulo 1 definimos un número de principios para usar registros administrativos con fines estadísticos. Los siguientes principios están relacionados con el sistema de registros:

<b>Principio de sistema</b>	Todos los registros estadísticos deberán incluirse en un sistema de registros coordinado. Este sistema garantizará que todos los datos puedan integrarse y usarse de manera efectiva.
<b>Principio de consistencia</b>	La consistencia de las poblaciones y de las variables es necesaria para la coherencia de los estimados de las diferentes encuestas por registros.

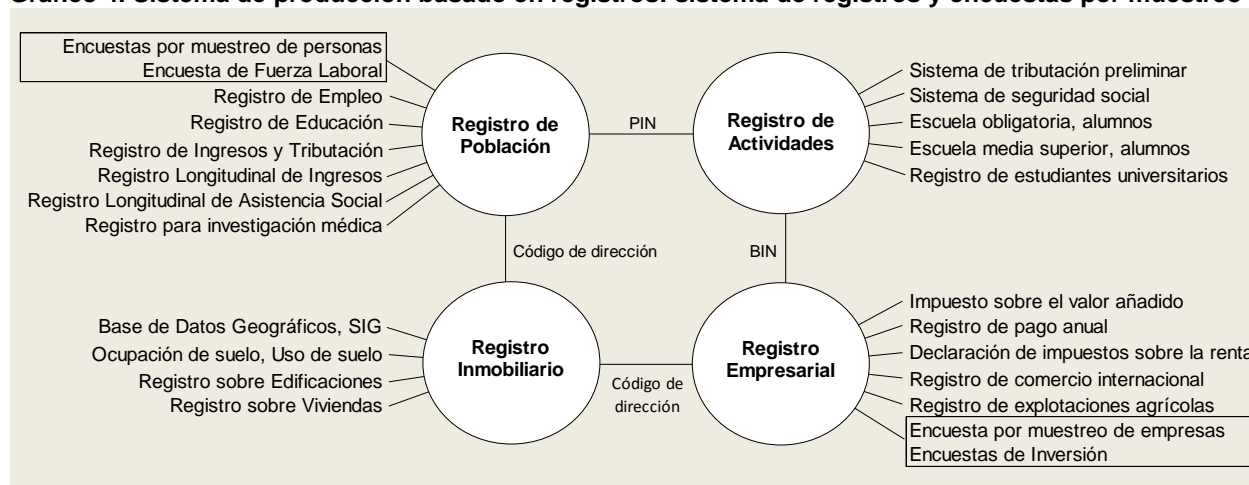
En un sistema estadístico nacional con un sistema coordinado de registros, es posible integrar todos los microdatos del sistema. En un sistema coordinado, las poblaciones de los diferentes registros son consistentes en lo relacionado a la cobertura, y las variables de los distintos registros no tienen valores contradictorios. Esto es lo que queremos decir con un *sistema estadístico integrado*. Por supuesto, es difícil conseguir la consistencia perfecta, pero si los problemas de cobertura y las inconsistencias en las variables son tolerables, podemos decir que el sistema es integrado. La integración de los microdatos será posible sin problemas perturbadores en lo que respecta a la falta de consistencia entre poblaciones y variables.



## ¿Cuándo y por qué deberíamos usar registros administrativos?

El Gráfico 4 ilustra un sistema de producción basado en registros. La base del *sistema de registros* consiste en cuatro registros base (círculos) que definen las poblaciones en el sistema y cierto número de registros estadísticos con variables estadísticas importantes. Las líneas en el gráfico muestran los vínculos importantes entre las unidades de los diferentes registros. También hemos añadido dos encuestas por muestreo: La Encuesta de Fuerza Laboral y la Encuesta de Inversión.

**Gráfico 4. Sistema de producción basado en registros: sistema de registros y encuestas por muestreo**



PIN = Número de identidad de las personas

BIN = Número de identidad de las empresas

Si las poblaciones y las variables son suficientemente consistentes, el sistema en el Gráfico 4 es un sistema integrado. Un sistema integrado tiene las siguientes características:

- Los microdatos de cualquier registro pueden combinarse con los microdatos de los demás registros. *Por ejemplo, empleo, educación e ingresos pueden combinarse 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 usar variables del Registro de Educación, en lugar de hacer preguntas acerca de esta temática. Así mejorará la calidad y reducirá los costos.*

En contraste con la integralidad y flexibilidad de los datos de los registros, los microdatos de la encuesta por muestreo no pueden combinarse con los microdatos de otras encuestas por muestreo. Esto se debe a que, por regla, únicamente un pequeño número de unidades se incluyen en más de una muestra, así que prácticamente no hay unidades en común que puedan combinarse.

- *Esto explica por qué no se necesita el enfoque de sistema en un sistema de producción tradicional sin registros, ya que es suficiente con considerar una encuesta por muestreo a la vez.*

Por lo anterior, las estadísticas basadas en registros y las estadísticas basadas en muestreo necesitan enfoques diferentes. El enfoque de sistema es fundamental para la estadística basada en registros; en contraste, las estadísticas basadas en muestreo no lo necesitan para nada. La bibliografía sobre teoría del muestro y metodología de encuestas por muestreo se concentran en tratar *una* encuesta por muestro, no un sistema de encuestas. Sin embargo, cuando creamos registros estadísticos debemos considerar simultáneamente varias fuentes a la vez, por lo que existen diferentes maneras de utilizar una misma fuente así como distintas maneras de utilizar las partes del sistema.

## ¿Cuándo y por qué deberíamos usar registros administrativos?

El Gráfico 5 muestra las unidades estadísticas y sus números de identidad en cuatro registros base. Con estas identidades, *todos* los registros pueden vincularse dentro del sistema. Puesto que los números de identidad de personas (PIN) o números de identidad de empresas (BIN) se incluyen en encuestas por muestreo, éstos también pueden vincularse al sistema. En el Registro de Ocupación, los números de identidad PIN y BIN se combinan. La ocupación es la unidad estadística y la misma persona puede tener varios trabajos y actividades estudiantiles durante un periodo específico. El Registro de Ocupación es el vínculo entre los Registros de Población y el de Negocios. Gracias a este vínculo, la estadística social y la estadística económica hacen parte del mismo sistema de producción.

**Gráfico 5. Los cuatro registros bases y sus vínculos**

Un sistema basado en registros también abre nuevas posibilidades para organizar el trabajo y eliminar la duplicación. 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 que alimentan el vector de ingresos dentro del sistema, etc. Así, estas *poblaciones y variables estandarizadas* se pueden utilizar en 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. Habrá *problemas de cobertura* si las poblaciones de los registros relacionados se comparan. Los registros en el Gráfico 4 pueden compararse de la siguiente manera:

- ¿Están incluidas todas las direcciones y residencias del Registro de Población y del Registro Empresarial el Registro Catastral?
- ¿Están incluidas todas las personas del Registro de Ocupación en el Registro de Población, y están incluidos todos los empleadores del Registro de Ocupación en el Registro Empresarial?

Si se comparan variables similares en los distintos registros, así como las variables de la encuesta por muestreo con variables similares en los registros, se encontrarán 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 (ISIC o CIIO) en distintas fuentes. Con el Registro de Ocupación será posible encontrar los empleadores de las personas entrevistadas en la Encuesta de Fuerza Laboral. La codificación de la actividad económica en la Encuesta de Fuerza Laboral puede entonces compararse con la codificación en el Registro Empresarial.

## ¿Cuándo y por qué deberíamos usar registros administrativos?

En resumen, hemos observado que 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 evaluación de la calidad.

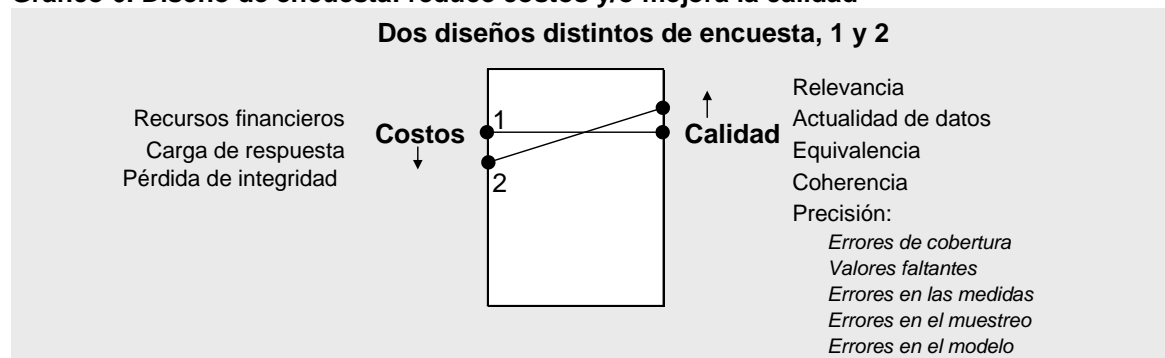
La creación de un sistema nacional integrado basado en registros requiere mucho trabajo y nuevas formas de pensamiento. Este proyecto de largo plazo necesita apoyo continuo de los directores y políticos de turno. La cooperación de los organismos gubernamentales y las autoridades administrativas también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

### 2.3 Comparación de costos y calidad

¿Por qué deberíamos usar registros administrativos con fines estadísticos? Hay dos razones que explican por qué los países deberían hacer la transición del modelo tradicional al modelo basado en registros: i) para reducir los costos, y ii) para mejorar la calidad. Cuando trabajamos en el diseño de encuestas, siempre buscamos conseguir estos dos objetivos.

En el Gráfico 6 se comparan dos diseños de encuestas. Si asumimos 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 ahorrar costos, ya que en el sistema basado en registros se pueden usar cuestionarios por correo o entrevistas por teléfono en lugar de enviar entrevistadores a visitar casas en una muestra por zona. Además, cuando las variables auxiliares del sistema de registros se usan para mejorar la estimación, la calidad puede mejorar ya que se reducirán los errores de muestreo y de no respuesta.

**Gráfico 6. Diseño de encuesta: reduce costos y/o mejora la calidad**



Se deberán reducir tres tipos de costos. Al pasar a un sistema basado en registros se pueden ahorrar *recursos financieros* y la *carga de respuesta*.

En muchos de los casos, la recopilación de datos molestará a los encuestados. Cuando la gente atiende una llamada telefónica del instituto estadístico por la noche, ésta puede verse como una *pérdida de integridad*. Después de una entrevista, el entrevistador puede haber obtenido información confidencial que podría ser una amenaza para la integridad de los encuestados. Podemos reducir estos costos con encuestas de registro.

En un sistema basado en registros, las identidades deberán ser reemplazadas por números anónimos para proteger la integridad y el acceso a los microdatos de los registros deberá estar restrin-

gido. Si no se toman estas medidas dentro del sistema basado en registros, los costos asociados con las pérdidas de integridad pueden aumentar.

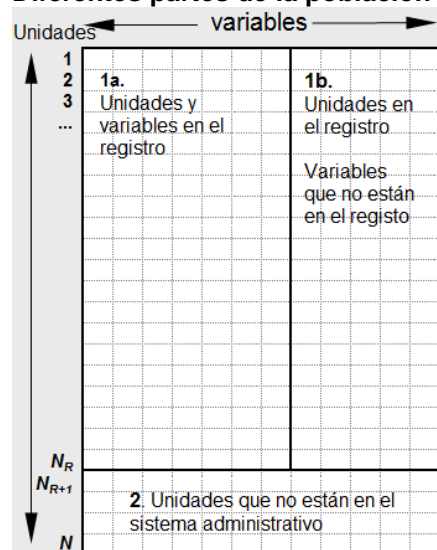
En cuestión de calidad, deberemos recordar que un sistema basado en registros abre nuevas posibilidades. En lugar de un censo tradicional de población y vivienda costoso, se pueden obtener un número de encuestas por registros que proporcionarán estadísticas anuales que no existían antes. Por ejemplo, las estadísticas anuales de pequeñas regiones sólo serán posibles en un sistema de producción basada en registros. También será posible crear registros longitudinales que serán muy importantes para los investigadores en diversas áreas temáticas. En estos casos, de hecho, comparamos una situación en la que faltan datos importantes de encuesta en un sistema tradicional con la calidad de las nuevas encuestas por registros.

## 2.4 Métodos combinados – muestreo por zona y por registros

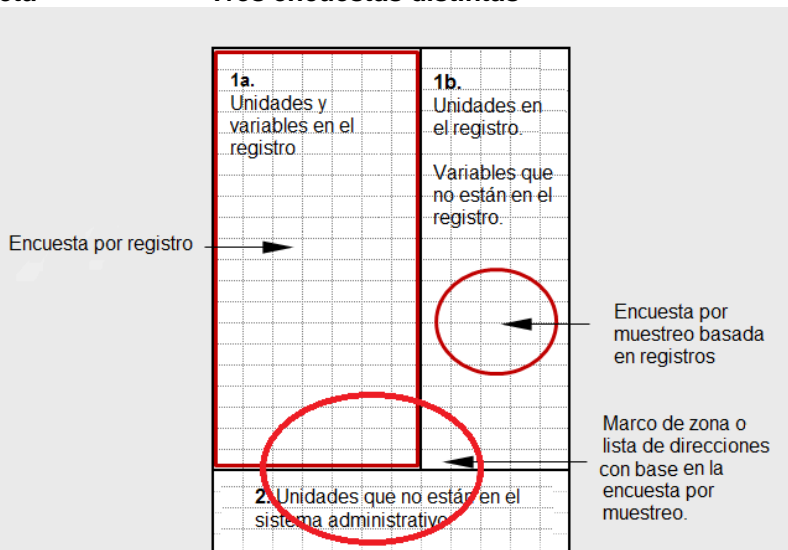
En un sistema basado en registros necesitamos tanto encuestas por registros como encuestas por muestreo como dos métodos complementarios. En Suecia, se realizan encuestas por muestreo cuando no se pueden encontrar variables importantes en fuentes administrativas y para estadísticas a corto plazo cuando 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 los campesinos no se encuentran incluidos en los registros administrativos. Por lo tanto, será necesario combinar encuestas basadas en registros y muestro basado en marcos de área. Asimismo, algunos sistemas administrativos no se actualizan de manera regular. En consecuencia, algunas variables importantes de los registros, como el domicilio, pueden estar incorrectos. La magnitud de estos errores se puede medir con encuestas por muestreo basadas en marcos de área.

**Gráfico 7.**  
**Diferentes partes de la población meta**



**Gráfico 8.**  
**Tres encuestas distintas**



Cuando un registro estadístico tiene errores de cobertura (sub cobertura o sobre cobertura), éste puede combinarse con una encuesta por muestreo basada en un marco de área. Dado que la muestra por zona en teoría no tiene errores de cobertura, la combinación de estimados combinados

## *¿Cuándo y por qué deberíamos usar registros administrativos?*

puede eliminar el sesgo. De igual manera, una encuesta por muestreo basada en registros con problemas de cobertura del marco del muestreo se puede combinar con una muestra por área.

En el Gráfico 7 la población meta está conformada de  $N$  unidades, y la población está dividida en dos partes:

- Parte **1** (**1a** y **1b**) formada de  $N_R$  unidades que se incluyen en el registro.
- Parte **2** formada por  $N - N_R$  unidades que conforman la cobertura insuficiente del registro.

La encuesta por registro en el Gráfico 8 puede combinarse con una muestra por área así como una muestra basada en registros puede combinarse con una muestra por área. En la Sección 5.3 explicamos cómo estas muestras por área pueden usarse para evaluación y ajuste de la calidad de los errores de cobertura.

### **2.5 Conclusiones: ¿cuándo y por qué deberíamos usar 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 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.

Si se usan los registros administrativos para la producción estadística, se pueden ahorrar costos y mejorar la calidad. Los censos pueden ser remplazados por encuestas por registros y se pueden desarrollar nuevos tipos de estadísticas regionales y longitudinales. El sistema nacional puede cambiarse a un sistema estadístico 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 zona.

### 3 ¿Cómo deberíamos usar los registros administrativos?

En Wallgren y Wallgren (2012) y (2014), explicamos los métodos que deberían usarse en el trabajo con registros administrativos. En esta sección damos una pequeña descripción de dichos métodos y destacamos aquellos que son nuevos o diferentes en un país que comienza a trabajar con el enfoque basado en registros.

En dichos países, la oficina nacional de estadística tuvo previamente acceso a las tablas estadísticas (macrodatos) producidas por los organismos gubernamentales y los institutos que son responsables de los registros administrativos. Un sistema estadístico integrado no puede crearse con macrodatos únicamente. Un sistema integrado deberá estar basado en registros estadísticos coordinados. El primer cambio importante tendrá que ver con que la oficina nacional de estadística tenga acceso a todos los *microdatos* que se requieren para desarrollar y mantener el sistema nacional de registros estadísticos. Usando únicamente los macrodatos es posible comparar las fuentes y encontrar errores e inconsistencias; no obstante, con los microdatos también es posible encontrar las causas de estos errores y hacer las correcciones para reducirlos.

#### 3.1 Creación de números de identidad para 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 seleccionada hará parte de la muestra, así como el mecanismo de entrevista que se implementará a dicha unidad. El nombre o identidad de una persona seleccionada en la muestra no es importante para la producción de estadísticas. Por lo tanto, la información con respecto a 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, esto es, en todos los censos, encuestas por muestreo y encuestas por registros. Los Registros de Población y los Registros de Negocios se usan para crear marcos de muestra, y las unidades estadísticas son identificadas 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 de muestra, las identidades se utilizan para contactar a las personas/establecimientos por correo o por teléfono. En las encuestas por registros, los microdatos tomados de los distintos registros se combinan para crear el conjunto de datos que se necesita para la encuesta en específico. Estas combinaciones se crean mediante vinculaciones de registros usando los números de identidad como claves de correspondencia.

En un sistema basado en registros hay tres aspectos que son importantes en relación a las identidades:

- La *calidad* de los números de identidad. Si la calidad es insuficiente, tendrá que haber un proyecto a largo plazo para mejorar esta parte en el sistema nacional de registro. A corto plazo, los números de identidad pueden suplirse con nombres, fechas de nacimiento y lugares de nacimiento.
- La *vinculación de registros* será una parte importante del proceso de producción cuando el sistema nacional de estadística se vuelva un sistema basado en registros. La razón es que todas las variables que se necesitan para una encuesta específica no aparecen en una sola fuente, por lo tanto las fuentes *deben* combinarse para crear las variables.
- El sistema de producción debe estar diseñado para *proteger la privacidad* usando números anónimos en lugar de números reales de identidad, y para garantizar 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 deberá ser buena. Los números de identidad nacional se usan para hacer que la administración de los sistemas tributarios y de seguridad social sea eficiente y cumpla con los derechos de los ciudadanos. Estos sistemas administrativos son dinámicos y regularmente sufren cambios y mejoras.

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 aportar parámetros generales de calidad 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 todos los recién nacidos y todos los nuevos inmigrantes, se evitan errores debido a cobertura insuficiente y duplicados. Los duplicados también pueden aparecer cuando las personas de edad vuelven a registrarse (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 hay una gran cantidad de datos administrativos que los utilizan. Se recurre a una *vinculación de registros determinística* si los números de identidad coinciden, lo cual implica un vínculo. Si los números de identidad no coinciden, no hay vínculo.

En un sistema basado en registros, siempre hay una vinculación al momento de crear el registro estadístico. Si, por ejemplo, cuatro fuentes se combinan para crear un registro estadístico, entonces se realizan cuatro operaciones de vinculación. Para algunos registros se hacen más de 100 operaciones de vinculación; ejemplos de éstos son los registros longitudinales, muy importantes en para la investigación temática. La conclusión de esto es clara: el índice de falta de correspondencia debe ser pequeño en cada operación de vínculo; de otra manera, 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 usan combinaciones de claves de correspondencias: nombres, fechas de nacimientos y lugares de nacimientos en lugar de números de identidad. Algunas veces estas variables aparecen registradas en formularios a mano. Puesto que los índices de falta de correspondencia no pueden ignorarse, este método no es adecuado para un sistema estadístico basado en registros. En un sistema basado en registros las *poblaciones* y las *identidades* deben ser manejables para que se pueda usar una vinculación de registros determinística.

- En Statistics Sweden, el Registro de Población contiene únicamente las personas registradas como residentes permanentes en Suecia. Todas estas personas tienen números de identidad válidos. Este principio afecta la cobertura, ya que hay personas que pertenecen a la población del censo deseado que no están formalmente registradas. A pesar de esta restricción, la ventaja de este sistema administrativo es que permite la vinculación de todos los registros a través de números de identidad válidos y disponibles para su uso.
- Cualquier país puede cambiar el sistema de números de identidad o, a través de sistemas paralelos, se les puede asignar números nuevos a las personas. Esta decisión debe tomarse

en la agencia que trabaja con los Registros de Población. Es posible que se requiera 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 puedan usar en la oficina de estadística.

### ***Protección de la privacidad***

Cuando un país cambia el sistema estadístico nacional tradicional a uno basado en registros, se deben desarrollar nuevos métodos que protejan la privacidad. Las identidades pueden y deben manejarse de tal manera 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. Supongamos que el formulario de impuestos de un contribuyente X deberá usarse 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 de la esposa de la persona X. Si estas personas son dueñas de una empresa, el formulario de impuestos de la empresa también debe revisarse, ya que puede haber transacciones sospechosas entre X, la esposa y la empresa. Una comparación con los formularios de impuestos del año anterior también debería hacerse.

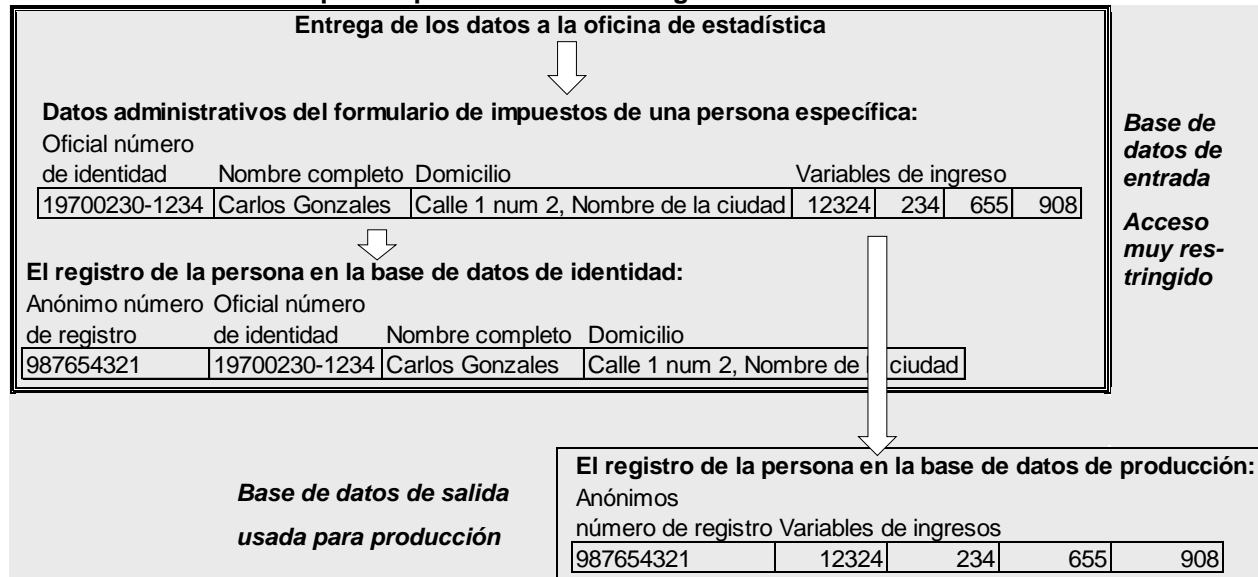
La situación descrita anteriormente es típica del trabajo con datos administrativos. Gracias a los sistemas de tecnologías de la información del sistema administrativo es fácil tener acceso a *todos* los datos pertinentes de una persona en específico, en este caso se combinan seis formularios de impuestos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

El trabajo en un instituto nacional se debe organizar de tal manera que el uso de números de identidad oficial y direcciones sea reducido y estrictamente restringido. De esta manera, la privacidad estará mejor protegida que antes. El proceso de este trabajo se ilustra en el Gráfico 9.

En un sistema nacional descentralizado, las estadísticas se producen en los organismos gubernamentales e institutos que son responsables del sistema administrativo. Estas estadísticas son generadas por funcionarios con acceso a los microdatos con nombres y direcciones fácilmente reconocidos. Cuando el sistema nacional se vuelve más centralizado y basado en registros, esta producción se reemplaza con el trabajo realizado en la oficina nacional de estadística *sin* acceso a nombres y direcciones.



**Gráfico 9. Un sistema TI para la protección de la integridad**



Una unidad en la oficina nacional de estadística consistente 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, los nombres y direcciones se eliminan y los números de identidad oficial se reemplazan por números anónimos. Después de que este trabajo está hecho, los datos anónimos se almacenan en la base de datos de salida y la base de datos de producción, la cual se utiliza para la producción de estadísticas. 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 entrada sólo incorporen información relacionada con la municipalidad o la región estadística pertinente en la que viva la persona.

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

¿Cómo usamos los datos obtenidos en las *encuestas por muestreo*? Después de la fase de recolección de datos tenemos *un* conjunto de microdatos. Este conjunto de datos se usará de *una* manera específica de acuerdo con los fines de la encuesta, bajo las premisas clásicas de la teoría de la probabilidad. Hay riesgo de que pensemos de manera similar cuando queremos usar registros administrativos para la producción estadística: tenemos acceso a *un* registro administrativo y tratamos de descubrir la *única* manera de usar este grupo de datos. Con esta manera de pensar tradicional no se utilizará todo el potencial de las nuevas fuentes con datos administrativos. En lugar de esto, el diseño de encuestas por registros deberá estar basado en los siguientes principios:

Definimos *una encuesta por registro* 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 diseño de una encuesta por muestreo:

- *un* registro administrativo se puede usar para *varias* encuestas por registro y registros estadísticos;
- para *una* encuesta por registro se necesitarán *varios* registros administrativos y registros estadísticos;

## ¿Cómo deberíamos usar los registros administrativos?

- el registro estadístico creado no será utilizado exclusivamente por su creador, sino también por sus colegas y por otras encuestas.

La encuesta de ingresos sueca basada en registros que se muestra en el Gráfico 10 ejemplifica lo anterior. Algunos registros administrativos externos son usados como datos de entrada. De la misma manera se usan algunos registros estadísticos internos obtenidos del sistema de registros de Statistics Sweden. Tres registros estadísticos son creados con estas fuentes; estos registros estadísticos constituyen los microdatos que son usados por la encuesta de ingresos. Algunas de las fuentes usadas como datos de entrada para la encuesta de ingresos también son usadas para otras encuestas por registro. Los formularios de impuestos obtenidos de la Oficina Tributaria Nacional y Liquidaciones de Ganancias también se utilizan para la generación de estadísticas económicas.

**Gráfico 10. Encuesta de ingresos sueca basada en registros**



### 3.3 Cuatro maneras de utilizar un registro administrativo

Cuando el instituto estadístico tiene acceso a nuevas fuentes de información, ¿cómo debería utilizar estos nuevos registros administrativos? ¿Cómo deberían modificarse las encuestas existentes? Para responder estas preguntas las fuentes administrativas deberán ser analizadas por especialistas temáticos y por metodólogos con un buen conocimiento general sobre el sistema de producción estadística.

Un registro o fuente administrativa puede utilizarse de cuatro maneras diferentes. La primera manera, la cual usa el registro administrativo únicamente, no aportará nada a un sistema integrado; pero las otras tres maneras mejorarán el sistema de producción, ya que lo vuelve más integrado.

#### 1. Completamente sola

Si la fuente tiene buena cobertura y las variables en la fuente son de buena calidad, la fuente puede utilizarse sola para producir estadísticas.

*Ejemplo:* En un sistema de registros de nacimientos, 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. Es claro que dicho registro recoge información de una sola 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.

#### 2. Sola, pero combinada con un registro base

El Registro de Población y el Registro de Empresas son dos registros bases 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 éste. 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.

*Ejemplo:* Si el Registro de Nacimientos y el Registro de Poblaciones pueden vincularse con la identidad de las madres, se puede utilizar el registro combinado para comparar las madres recientes con todas las mujeres que no lo son y producir estadísticas sobre el *número de nacimientos por cada 100 mujeres* por región y por edad de la madre.

### **3. Combinada 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. Así que 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.

*Ejemplo:* Si se pueden vincular el Registro de Nacimientos, el Registro de Población y el Registro 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 1000 mujeres por región, edad y *nivel de educación* de la madre.

### **4. Para mejorar otras encuestas, es decir 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, será también posible corregirlos o reducirlos. Estos errores los tratamos en Wallgren y Wallgren (2015a). También explicamos cómo encontrar y reducir errores e inconsistencias entre las variables de clasificación que se encontraron cuando se comparó la actividad y el sector económico en la Encuesta de Fuerza Laboral Sueca y la encuesta de Estadísticas Estructurales de Empresas.

*Ejemplo:* El Registro de Nacimiento mencionado antes puede utilizarse para mejorar la calidad del Registro de Población en tres formas:

- La información sobre la dirección relacionada con madres recientes en el Registro de Nacimiento se puede usar para actualizar la dirección en el Registro de Población.
- Es posible que algunas madres no estén en el Registro de Población. Esta cobertura insuficiente puede reducirse si los datos del Registro de Nacimientos de estas madres se añaden al Registro de Población.
- Además, se pueden incluir los bebés recién nacidos al Registro de Población estadístico.

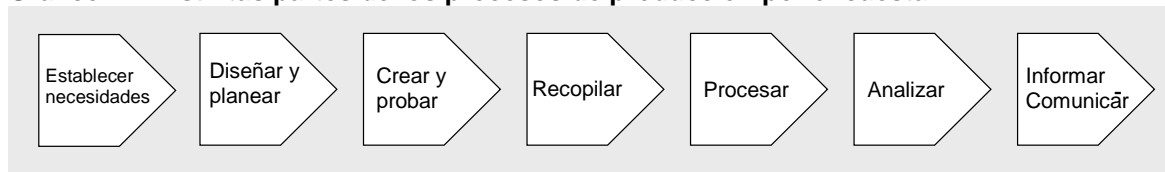
Todas las encuestas por registro en países con sistemas estadísticos basados en registros (por ejemplo, los países nórdicos) utilizan un registro base con el método 2 y 3 mencionados anteriormente. Los países con nuevos registros que no ha desarrollado aún registros bases de buena calidad deberán empezar con encuestas por registros de tipo simple, como el mencionado en el método 1. Cuando ya se han desarrollado registros bases, será posible crear encuestas por registros de acuerdo con los métodos 2 y 3. Hasta ahora, el método 4 no ha sido practicado ni en los países nórdicos.

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

El Modelo Genérico de Proceso Estadístico para Empresas (MGPEE) ha sido discutido en algunas oficinas nacionales de estadística. El foco de este modelo está en las distintas partes del proceso de producción para *una* encuesta. El Gráfico 11 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 de tecnologías de la información comunes.

**Gráfico 11. Distintas partes de los procesos de producción por encuesta**



El modelo del proceso es aplicable a encuestas por muestreo y censos; pero no resulta de ayuda para el enfoque de sistemas que es esencial en las encuestas por registro. El trabajo esencial relacionado con la creación, mantenimiento y utilización del sistema de registros no se incluye en el modelo del proceso. El modelo muestra que el enfoque de sistemas no es necesario en un sistema de producción tradicional sin registros, ya que considerar una encuesta por muestreo a la vez es suficiente.

En contraste con la filosofía que subyace al Modelo Genérico del Proceso Estadístico para Empresas, el modelo de sistemas busca obtener *variables y poblaciones estandarizadas*, para que las encuestas se vuelvan consistentes y coherentes. 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, dando la oportunidad de encontrar y reducir errores. Los Gráficos 12 y 13 en la Sección 4.1 muestran los modelos de proceso adecuados para encuestas por registro.

## 4 ¿Cómo hacer la transición del sistema antiguo al nuevo?

Un sistema estadístico tradicional basado en el muestreo de áreas puede ser más o menos integrado. 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 son distintos de los datos de la encuesta por muestreo, por lo cual aparecen ahora otro tipo de inconsistencias. En efecto, los registros presentan distintas coberturas, valores de variables en conflicto y variables similares con distintas definiciones. No obstante, a diferencia del método por muestreo, ahora si es posible combinar y comparar los microdatos que provienen de las distintas fuentes, y, de esta forma, los errores e inconsistencias se pueden corregir. A través de este proceso, el cual demandará mucho trabajo y requerirá resolver varios problemas, se empezará a desarrollar gradualmente un sistema integrado de registros con poblaciones y variables consistentes.

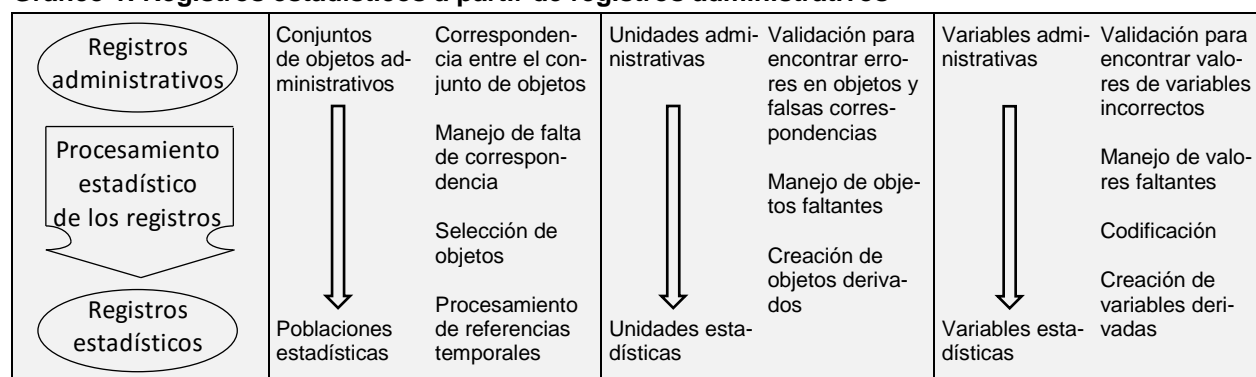
### ¿Cómo organizar el trabajo cuando se tiene acceso a los registros administrativos?

Utilizamos dos modelos diferentes cuando abordamos este asunto. El *modelo de proceso de producción* en el Gráfico 12 (véase la página siguiente) es la base para describir cómo organizar el trabajo en la oficina de estadística. El *modelo de sistema de registros* (ver Gráfico 4) es la base para describir cómo utilizar los registros administrativos para desarrollar un sistema integrado de registros estadísticos. Los asuntos metodológicos relacionados con el proceso de producción se tratan de manera más detallada en la Sección 5.

### 4.1 Modelo de proceso de producción

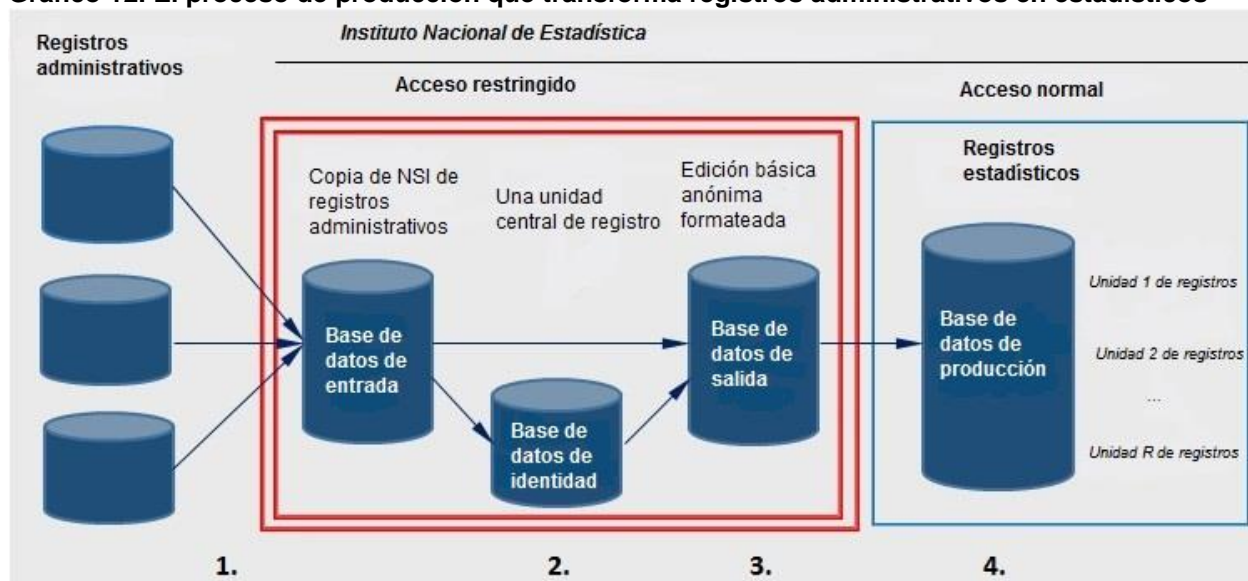
El proceso de transformación de registros administrativos a registros estadísticos se muestra en el Gráfico 1, en donde se destacan los *métodos estadísticos*.

**Gráfico 1. Registros estadísticos a partir de registros administrativos**



El mismo proceso también se ilustra en el Gráfico 12, en el que se muestra la *organización del trabajo*. Como se puede ver, en la parte inferior del Gráfico se estructura el proceso de producción en cuatro pasos numerados del 1 al 4.

**Gráfico 12. El proceso de producción que transforma registros administrativos en estadísticos<sup>2</sup>**



1. Las autoridades administrativas entregan los registros administrativos a la oficina de estadísticas. La selección correcta de unidades y variables que se entregarán debe realizarse buscando que cubra los requisitos estadísticos. La cooperación entre las autoridades administrativas y la oficina de estadísticas deberá estar bien definida y ser permanente. Para la actualización del Registro de Población y el Registro de Empresas, se deberán entregar únicamente los registros nuevos y los cambiados.
2. Con la Base de Datos de Identificación, las variables identificadoras como número de cédula, otros tipos de número de identidad y nombres son remplazadas por números anónimos de registro. A la base de datos de salida se transmiten únicamente estos números de registro. Los registros con distintos números de cédula para la misma persona se vinculan con el mismo número anónimo de registro.
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 los estándares y formatos usados en la oficina de estadística. Además, se realiza una validación básica de las variables en cada registro administrativo en la base de datos de salida.
4. El trabajo central de la creación de registros estadísticos se realiza en la base de datos de producción. Se combinan diferentes fuentes mediante vinculación de registros, se crean poblaciones de registros, se crean unidades derivadas y variables derivadas, y se lleva a cabo una validación para consistencia. Esta parte del proceso se muestra en el Gráfico 1 en la Sección 1.3.

### **Contactos con las autoridades administrativas**

Los contactos con las autoridades administrativas son necesarios para la oficina de estadística por diversos motivos. La oficina de estadística no solo necesita los microdatos, sino también los metadatos con definiciones de todas las variables administrativas. El proceso administrativo que genera los registros administrativos debe comprender: -¿cómo se generan el conjunto de objetos administrativos?-. Estos conjuntos de objetos serán usados para crear poblaciones estadísticas.

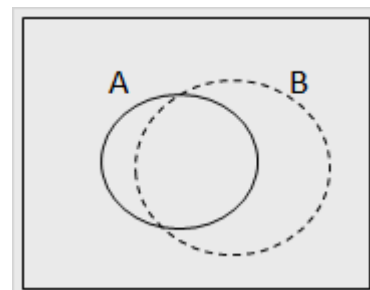
<sup>2</sup> Este modelo se desarrolló junto con David Chaín, INEC, durante nuestra visita a Ecuador en 2015.

## *¿Cómo debemos hacer la transición del sistema antiguo al nuevo?*

Por lo tanto, los estadísticos deben conocer el sistema administrativo para poder evaluar la pertinencia de los datos.

Los contactos con la oficina de estadística 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 con el sistema de registros estadísticos. La falta de calidad estadística podrá indicar también falta de calidad administrativa. La cobertura insuficiente no es un problema estadístico únicamente, de igual manera afecta el proceso administrativo. Las duplicaciones o domicilios erróneos también son problemas administrativos.

La selección de unidades y variables que serán enviadas a la oficina de estadística debe realizarse de manera que cumpla con los requisitos estadísticos. La autoridad administrativa a menudo tiene mucho más datos de los que esperan los estadísticos. El rectángulo de la derecha describe los datos que la autoridad administrativa tiene, el área A describe los datos que la autoridad piensa que los estadísticos necesitan y el área B punteada describe los datos que el instituto de estadística debería tener. Para evitar malos entendidos, la selección de los datos que se enviarán deberá decidirse mediante una plática entre las dos partes.



La cooperación con las autoridades administrativas deberá manejarse con grupos de trabajo de la oficina de estadística. Estos grupos de trabajo estarán integrados por el personal con la competencia en el asunto relacionado, en métodos estadísticos y de registros, así como en tecnologías de la información. La cooperación entre estadísticos y los distintos organismos gubernamentales y autoridades administrativas deberá ser un proyecto a largo plazo, con el objetivo de desarrollar sistemas nacionales estadísticos y administrativos. En la Sección 5.1 se trata el asunto relacionado con el Registro Civil y el sistema de estadística.

### ***La unidad central de registro***

La Base de Datos de Identidad ha sido desarrollada con dos fines – proteger la integridad y facilitar el proceso de registro. La integridad será protegida ya que los nombres y números de cédulas serán eliminados – el personal que trabaja con las estadísticas basadas en el registro únicamente tendrá acceso a números de registro anónimos cuando se generen 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 y se reducirá el riesgo de errores. Esto debido a que todos los datos utilizados en la Base de Datos de Producción se ha editado de tal forma que las variables de identificación en conflicto como nombres y diferentes tipos de números de identidad han sido reemplazadas por un sistema unificado de números de registro que son fáciles de utilizar.

Una unidad central con un número limitado de personal debe organizarse para ser responsable de recibir todos los registros administrativos. Únicamente estas personas deberán 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 reemplazar 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 los enlaces entre identidades y números de registro anónimos están almacenados.

## *¿Cómo debemos hacer la transición del sistema antiguo al nuevo?*

La unidad central también será responsable del manejo de las identidades utilizadas en las encuestas por muestreo basadas en los registros. Una encuesta por muestreo basada en registros se debe llevar a cabo de la siguiente manera:

1. Se toma una muestra de personas al azar de un marco creado con el Registro de la Población. Como resultado se obtiene una lista con los números de registro anónimos para la muestra.
2. Esta lista se envía a la unidad central. Ésta reemplaza los números de registro con los nombres y direcciones que se utilizan 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.*
3. Los datos recolectados son entregados a la unidad central, la cual elimina los nombres y direcciones y los reemplaza con los números de registro anónimos.
4. El conjunto de datos anónimos es entregado 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 estándares que deben utilizarse en la oficina de estadística.  
*Si por ejemplo, tres diferentes fuentes administrativas en la Base de Datos de Entrada definen sexo de diferentes maneras: H/M, 0/1 ó 1/2 para hombres/mujeres. Entonces en la Base de Datos de Salida, todos los registros deben utilizar los mismo códigos, ej. 1 para hombres y 2 para mujeres.*  
Este tipo de estandarización será mucho más importante cuando el sistema de producción cambia de estar basado en un marco de área a un sistema basado en registro, ya que se podrán combinar microdatos de diferentes registros y encuestas y esos mismos formatos y estándares se podrán utilizar en todos los registros y encuestas.
- Validación básica de cada registro administrativo. Esta validación consiste en dos tareas: buscar duplicados y encontrar y corregir valores erróneos de variables. Esto está descrito en Wallgren y Wallgren (2014, Sección 9.1.1). El manejo de duplicados también se comenta en la Sección 5.2 más adelante.

### ***¡Hazlo correctamente desde el principio!***

Las Bases de Datos de Identidad y los estándares y formatos que deben utilizarse para las variables del sistema deben establecerse al principio del proceso de modernización del sistema estadístico. Si no se sigue esta regla, se tendrán que desarrollar sistemas de tecnologías de la información (TI) costosos que utilicen identidades y formatos 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 de TI será muy costoso.

### ***Bases de Datos de Producción***

Los registros estadísticos son creados por diferentes unidades de registro en las Bases de Datos de Producción. Cada unidad puede importar los registros administrativos que necesita desde la Base de Datos de Salida. También se pueden utilizar los registros estadísticos del sistema – todos los registros del sistema 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 la vinculación de regis-



tros. Estas poblaciones, unidades y variables son revisadas para ser editadas consistentemente, y este trabajo de validación trae como resultado el descubrimiento y la reducción de errores e inconsistencias.

**Gráfico 13. Los dos tipos de validación en el trabajo de creación de registros estadísticos**



Existen dos tipos diferentes de validación en el trabajo para crear registros estadísticos:

- El primer tipo de validación (cuadro punteado) la realiza la unidad central de registro cuando edita cada registro administrativo en la Base de Datos de Salida. Después de buscar los duplicados, se realiza la validación con los mismos métodos que se utilizan para editar encuestas por muestreo. El objetivo es encontrar y corregir los valores de variables erróneos.
- El Segundo tipo de validación (elipse continua) es la validación de consistencia que llevan a cabo las unidades responsables de cada registro estadístico en la Base de Datos de Producción. Se encontrarán muchos tipos de errores e inconsistencias diferentes al comparar las distintas fuentes. Además de encontrar valores de variables erróneos, también se encontrarán errores de cobertura y errores relacionados con las unidades estadísticas.

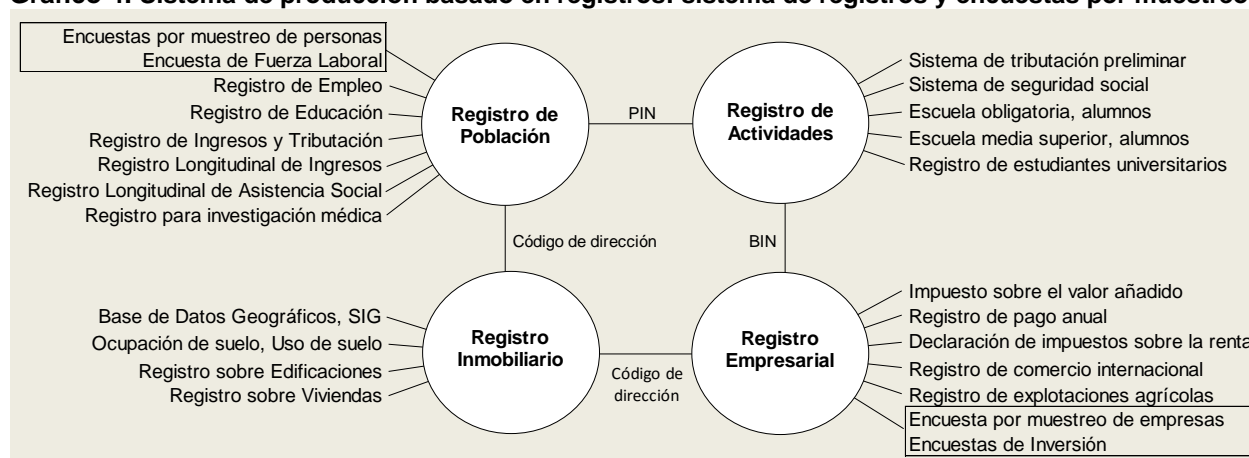
El trabajo de validación que se realiza en la Base de Datos de Salida y en la Base de Datos de Producción es la parte central del trabajo con el aseguramiento de calidad de las estadísticas basadas en registros. La documentación de los resultados del trabajo de validación es importante – esto hará posible monitorear cómo cambia la calidad de los datos administrativos con el tiempo. El aseguramiento de la calidad se comenta en la Sección 4.3.

## 4.2 El modelo de sistema de registro

Los registros estadísticos se crean en la Base de Datos de Producción. En el marco del modelo de sistema de registro se comentan los siguientes temas:

- cómo organizar el trabajo en relación con la competencia del objeto,
- cómo organizar el trabajo para lograr la consistencia,
- cómo llevar a cabo la vinculación de registros, y
- cómo utilizar el sistema para diseño de encuestas.

**Gráfico 4. Sistema de producción basado en registros: sistema de registros y encuestas por muestreo**



PIN = Número de identidad de las personas

BIN = Número de identidad de las empresas

### Competencia del objeto

El modelo del sistema está estructurado por tipo de unidad estadística: personas, actividades, empresas y bienes raíces o terrenos. Esta estructura está relacionada con diferentes campos de objetos. Por ejemplo, la estadística económica está relacionada 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 del objeto? Tomemos los cuatro registros con datos sobre personas como ejemplo: los registros de Población, Empleo, Educación e Ingreso e Impuestos. Los registros anuales se crean cada año: la población de personas al 31 de diciembre. Estos datos se pueden estructurar en cuatro registros diferentes o en un registro integral de estadística social:

**Gráfico 14. ¿Cuatro registros o un registro integral?**

Registro de Población	Registro de Empleo	Registro de Educación	Registro de Ingresos e Impuestos	Registro de Estadística Social
PIN Edad Sexo Municipio donde vive	PIN Edad Sexo Municipio donde vive Estado de empleo ISIC o CIU Municipio donde trabaja	PIN Edad Sexo Municipio donde vive Nivel más alto de ed. Tipo de educación	PIN Edad Sexo Municipio donde vive Ingresos por trabajo Ingresos por ... Ingresos disponibles	PIN Edad Sexo Municipio donde vive Estado de empleo ISIC o CIU Municipio donde trabaja Nivel más alto de ed. Tipo de educación Ingresos por trabajo Ingresos por ... Ingresos disponibles

## *¿Cómo debemos hacer la transición del sistema antiguo al nuevo?*

Aquí hay dos preguntas: ¿Cómo se debe organizar el trabajo y cómo se deben almacenar los datos? Nuestra respuesta a la primera pregunta es que el trabajo se debe organizar para fortalecer el papel de la competencia del objeto. 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, del Registro de Empleo y el Registro de Seguridad 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 de objeto; los datos son entregados por diferentes autoridades administrativas y tiene diferentes tipos de usuarios. Por lo tanto es preferible organizar cuatro unidades diferentes de registro cuando se haya establecido la producción de estadísticas.

Nuestra respuesta a la segunda pregunta es que se puede elegir almacenar los datos en cuatro registros independientes y aceptar alguna redundancia, o almacenar los datos en un registro integral sin ninguna redundancia. Con base en la experiencia de Statistics Sweden, 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 esta sección, resaltamos la importancia de la competencia del objeto. Cuando el sistema estadístico se convierte en uno basado en registros, el papel de este tipo de competencia cambia. Los contactos con autoridades administrativas y la comprensión de los sistemas administrativos nacionales son áreas nuevas e importantes donde se necesita esta competencia. ¿Cómo deberían diseñarse los nuevos registros estadísticos para proporcionarnos información importante sobre la sociedad? Aquí se necesita competencia de objeto.

El sistema de registro está estructurado por poblaciones y variables – realmente por el contenido de los registros. Esto significa que las diferentes partes del sistema de registro están vinculadas a los diferentes tipos de competencia de objeto. El trabajo con registros estadísticos se debe organizar para promover esta competencia de objeto.

Durante la década de los 90, se crearon muchos registros estadísticos nuevos en Statistics Sweden. El trabajo fue dirigido por especialistas en objetos. Las unidades de registro podían contactar a las autoridades y a los futuros usuarios de las nuevas estadísticas y utilizar su creatividad para desarrollar nuevos registros. Ésta era una organización descentralizada en la cual los emprendedores podían 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. Los gerentes se deben involucrar en esto – cuando se en-

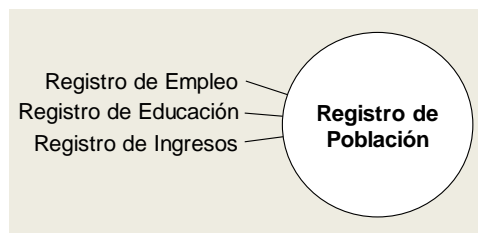
## ¿Cómo debemos hacer la transición del sistema antiguo al nuevo?

cuentran errores e inconsistencias se deben tomar decisiones, en relación con cuáles registros deben ser corregidos o cambiados, y cómo deben asignarse 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 y otra unidad con estadísticas anuales, es posible olvidar esta coordinación debido a la diferencia en tiempo. Supervisar y fortalecer este trabajo en equipo será una tarea importante para los gerentes. A continuación se mencionan dos ejemplos de este tipo de cooperación:

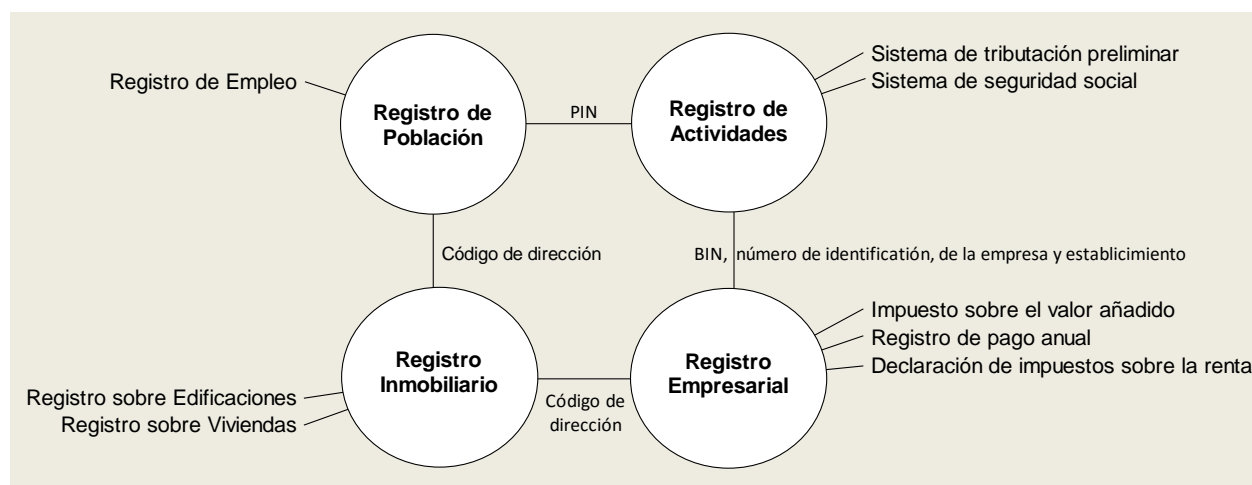
### 1. Cuatro registros anuales

En primer lugar, 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 sobre empleo, educación e ingresos. De esta sencilla manera, logramos una coordinación perfecta con respecto a poblaciones y variables.



### 2. Registro de empleo

El registro requiere cooperación con políticos y cooperación entre diferentes departamentos en la oficina nacional de estadística.



Primero, los patrones con más de un establecimiento deben informar en cuál establecimiento trabaja cada empleado. Puede ser necesaria una decisión política para que esta información sea parte del sistema de seguridad social o fiscal.

Segundo, se necesitan los 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 para personas, empresas, establecimientos y códigos postales para la localización de los hogares de las personas y la localización de los establecimientos donde trabajan las personas. La unidad responsable del Registro de Empresas debe crear un registro de todos los establecimientos. El objetivo principal de este registro es apoyar al Registro de Empleo. Se necesita cooperación para lograr consistencia en poblaciones – la población de empleados y de personas independientes debe ser consistente con el Registro de Población y la población de patrones y personas independientes debe ser consistente con el Registro de Empresas.

### ***Vinculación de registros y diseño de encuestas***

Cuando se discute la vinculación de registros en las publicaciones científicas, generalmente se refiere a la vinculación de los registros entre personas. En un sistema basado en registros, la vinculación de registros es más avanzada y brinda más posibilidades, ya que se puede vincular por ejemplo personas con empresas o viviendas, y actividades con personas o empresas.

Esta característica de un sistema integrado crea nuevas oportunidades para el diseño de encuestas. El Registro de Empleo 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 Statistics Denmark (Stender et al, 2015). Un gran volumen de datos sobre actividades de trabajo, estudio, así como la participación en diferentes programas de bienestar social se utiliza para describir la participación de la población en el mercado laboral. Se requiere la competencia del objeto para crear registros de este tipo, por lo que es importante trabajar junto con los investigadores académicos. El análisis del objeto define el contenido deseado; para luego llevar a cabo el procesamiento avanzado de registros de muchas fuentes para crear el registro deseado.

### **4.3 Aseguramiento de la Calidad**

El registro estadístico de población es la parte más importante de un sistema de producción basado en registros. Sin este registro base, será difícil crear estadísticas basadas en registros; por el contrario, si este registro existe, será más fácil continuar y desarrollar otras partes del sistema. En este sentido, cuando un país ha decidido desarrollar estadísticas basadas en registros, es importante que lleve a cabo con prioridad el desarrollo de un registro estadístico de población. Una vez creada la primera versión de este registro, el siguiente paso es evaluar el registro. Las encuestas por muestreo existentes basadas en marcos de área se pueden utilizar con este fin.

Los registros administrativos de población en los países latinoamericanos que visitamos tienen dos fallas principales: una cobertura insuficiente (subregistro), y domicilios residenciales incorrectos como consecuencia de su desactualización. Esto se comentó en la sección 2.4 y se continuará comentando en la sección 5.3, en donde se describe la importancia del muestreo por áreas para estimar estos errores.

Otro tipo de aseguramiento de la calidad es el trabajo en el que se analiza la utilidad de cada registro administrativo. Wallgren y Wallgren (2014, pp. 129, 176, 273-277) describen cuatro grupos de indicadores de calidad que se pueden utilizar con este fin.

El primer grupo de indicadores mide la *relevancia*, y permite analizar la calidad de los metadatos en relación con cada registro administrativo. Esta información consiste en la documentación creada por la autoridad administrativa responsable de la fuente. Las formas administrativas que proveen instrucciones y entrevistas con los proveedores de datos deben utilizarse para este análisis. La relevancia de la población, las unidades y las variables administrativas son evaluadas con base en estos metadatos.

El segundo grupo de indicadores mide la *precisión*, el cual se basa en el análisis de los microdatos de cualquier registro administrativo. La calidad de las claves de concordancia, la existencia de duplicados, los valores faltantes y los valores incorrectos están documentados.

El tercero y cuarto grupos de indicadores se basan en comparaciones con otras fuentes y registros del sistema. Las cuestiones de cobertura e inconsistencia entre variables se analizan con estos indicadores. Laitila et al (2012) presentan estos cuatro tipos de indicadores; y un caso práctico muestra cómo se pueden utilizar para evaluar la calidad de un registro administrativo.

## **5 Problemas relacionados con el Registro de Población**

La calidad del registro estadístico de población depende de la calidad de los registros administrativos utilizados como entrada. Los registros administrativos de baja calidad pueden ser compensados hasta cierto punto con los métodos estadísticos adecuados que permiten combinar la información de otros registros.

Wallgren y Wallgren (2014, Sección 5.3) contiene una descripción general de variables que deben incluirse en el Registro estadístico de Población. También se ilustra cómo se actualiza este registro con nueva información. Las variables de identificación, son las variables de referencia con vínculos a otros tipos de unidades estadísticas y referencias de tiempo<sup>3</sup> con fechas de eventos demográficos, estas variables son importantes dentro del Registro de Población.

Esta sección, primero aborda el *sistema nacional de registro* el cual determina cómo son registradas las personas en el Registro Civil<sup>4</sup>. ¿Cómo se puede mejorar el sistema de registro? Creando un sistema estadístico eficiente basado en registros puede requerir mejorar las rutinas utilizadas para crear y mantener el Registro Civil.

En segundo lugar se discute cómo encontrar y tratar los *duplicados*. Una persona puede tener registros en diferentes momentos, por lo que esta información debe mantenerse en el sistema de registro de población para poder rastrear los cambios en el tiempo.

Finalmente, esta sección aborda el tema de la *cobertura insuficiente* en el registro de población y el problema con los *domicilios residenciales* desactualizados. Para monitorear estos problemas de calidad, es necesario llevar a cabo una encuesta por muestreo basada en marcos de área. Tal vez sea posible hacer ajustes por este tipo de errores con los mismos métodos utilizados para corregir las estimaciones por falta de respuesta.

### **5.1 El Registro Civil y el sistema nacional de estadística**

La Oficina de Registro es la autoridad responsable de registrar a las personas, así como 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 estadístico de población será el registro más importante en un sistema estadístico nacional moderno. Cuando las estadísticas sociales cambian a estar basadas en registros, puede ser necesario mejorar las rutinas empleadas por la Oficina de Registro – de lo contrario, no será posible crear estadísticas basadas en registros de suficiente calidad. La Oficina de Registro y el instituto de estadística deben trabajar juntos en un proyecto a largo plazo con este fin.

En los países nórdicos, los registros administrativos de población y los sistemas de números de identidad fueron de buena calidad desde el principio. Cuando los países nórdicos empezaron a producir estadísticas basadas en registros en los 60, podían utilizar una sola fuente para el registro de población, ya que la calidad de esa fuente era buena. Además, el sistema de números de iden-

---

<sup>3</sup> Recomendamos que las referencias de tiempo tengan el formato AAAAMMDD. Las variables de tiempo definidas de esta forma son variables de escala ordinaria y pueden ser utilizadas fácilmente para definir poblaciones de diferentes registros.

<sup>4</sup> En todos los países, una autoridad es responsable del registro administrativo de población. Aquí llamamos registro al *Registro Civil* y la autoridad a la Oficina de Registro. Utilizamos el término *número de cédula* al número de identificación personal (PIN).

tividad estaba bien establecido y fue fácil de usar desde el principio. Para los países nórdicos fue posible utilizar métodos sencillos para obtener estadísticas sobre personas basadas en registros.

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 residenciales del Registro Civil pueden tener una baja calidad, lo que significa que la información sobre domicilios en otras fuentes debe utilizarse para el registro estadístico de población. Se pueden tener en uso diferentes sistemas paralelos para identidades y se pueden 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 DANE en Colombia ha desarrollado un método para clasificar a las personas como emigrantes o inmigrantes. Los pasaportes de las personas que salen o entran a Colombia son escaneados y clasificadas como personas que han salido o entrado permanentemente del país con base en los datos administrativos de un periodo de al menos 12 meses. Si se combinan estos datos con los números de identidad, se pueden utilizar para actualizar el registro estadístico de población. Se recomienda que este método también se utilice en otros países<sup>5</sup>. De todas formas, la mejor solución 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, se ha discutido recientemente cómo se puede mejorar el registro nacional, sobre la mesa hay dos alternativas:

- Hacer que la Oficina de Registro se vuelva responsable de actualizar los domicilios residenciales y la información sobre la composición de los hogares.
- Hacer que los municipios sean los responsables de crear y mantener los registros municipales con los domicilios residenciales y la composición de los hogares actualizados.

Sin embargo, es de tener en cuenta que la alternativa de que los registros sean manejados por los municipios tomara un mayor tiempo de implementación. Por lo tanto, la Oficina de Registro debe ser la responsable de actualizar los domicilios en el Registro Civil mientras que el Instituto del Seguro Social será el responsable de actualizar la composición de los hogares.

Se discutió una tercera alternativa durante nuestra visita en 2015. El instituto nacional de estadística podría ser el responsable de una plataforma de Internet en la cual las instituciones como la Secretaría de Educación, las universidades, el Instituto del Seguro Social, la Autoridad Fiscal, etc., fueran responsables de actualizar la información sobre la población. Cuando los ciudadanos necesiten el servicio de estas instituciones, ellos mismos deberán proporcionar información actualizada a través de la plataforma de Internet – ¿dónde viven y con quién?

---

<sup>5</sup> Es más, Suecia podría utilizar el mismo método, ya que existe exceso de cobertura en el Registro de Población Sueco, el cual se puede reducir con este método.

Nuestra recomendación se basa en nuestra experiencia en Suecia. En lugar de crear estructuras paralelas, cada una trabajando con datos que son relevantes para el sistema nacional de registro, la Oficina de Registro debería tener toda la responsabilidad. Los municipios y otras autoridades deben cooperar y apoyar a la Oficina de Registro en esta tarea. Debe ser obligatorio para estas autoridades presentar toda la información y las correcciones solicitadas por la Oficina de Registro en la tarea para mantener y actualizar el Registro Civil. En Suecia, de esta forma se presentan alrededor de 1.2 millones de correcciones en la Oficina de Registro cada año.

El instituto nacional de estadística debe participar en este trabajo de mejora del sistema de registro y puede contribuir midiendo la calidad y las mejoras en calidad. Sin embargo, siempre debe seguirse el principio de *tráfico en un solo sentido*: los microdatos pueden ir de las autoridades administrativas al instituto de estadística; pero nunca en la dirección opuesta.

### ***La utilización de los medidores de energía eléctrica como marco de registro***

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

Si casi todos los hogares son usuarios de energía eléctrica, entonces el registro de usuarios de las compañías de electricidad se puede utilizar como un marco de hogares. Este marco se podría utilizar para generar encuestas por muestreo basadas en registros que podrían ser una alternativa rentable al muestreo basado en marcos de área. El registro de medidores de energía eléctrica también se podría utilizar como marco para remplazar la información de la vivienda en los censos de vivienda por un cuestionario enviado a todos los hogares o a una muestra de los hogares.

### ***¿Cómo se deben entregar los datos al instituto de estadística?***

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. Se han planeado actualizaciones trimestrales. La entrega consistió en dos registros, un registro con las personas que se habían registrado y recibido un número de cédula, y un segundo registro que consistió de las personas con acta de nacimiento pero sin registro con un número de cédula.

La entrega de datos al INEC es una nueva tarea para la Oficina de Registro, la cual es susceptible de mejora. Sólo hubo un registro para cada número de cédula, por lo cual sospechamos que la población activa al 31 de diciembre fue la única entregada. Esto significa que no es posible crear estadísticas de población hacia atrás en el tiempo.

El INEC ha realizado un trabajo cuidadoso al editar los datos del Registro Civil y ha informado sobre la estructura de los errores al personal de la Oficina de Registro – su primera experiencia en retroalimentación sobre calidad de expertos externos.

Cuando el INEC recibió la primera actualización después del primer trimestre de 2015 surgió otro problema – la Oficina de Registro nuevamente sólo pudo enviar la población de alrededor de 18 millones de registros, no solo los registros cambiados.

Estas experiencias dejan ver la necesidad de poner toda la atención a esta parte del sistema nacional de estadística. Entregar datos al instituto de estadística es una nueva tarea para las autoridades administrativas – ¿qué tipo de datos se necesitan y cómo se deben seleccionar y procesar los datos para entrega? ¿Cómo debe utilizar la nueva retroalimentación de calidad la autoridad administrativa?



## 5.2 Cómo encontrar duplicados

En el INEC, el problema de los duplicados en el Registro de Población se considera un problema importante. Se pueden encontrar los duplicados comparando los nombres, fechas de nacimiento y lugares de nacimiento – la misma persona se ha registrado dos veces y tiene dos números de cédula o un registro con número de cédula y otro sin. El INEC sabía que los registros que recibía de la Oficina de Registro contenían duplicados. Pero ¿cómo encontrar estos duplicados y cómo manejarlos? Se sugiere la solución descrita a continuación.

Los registros de población pueden ser un arduo proceso. El INEC trabaja con registros que consisten en 18 millones de registros. Si se comparan 18 millones de registros con todos los demás registros esto involucraría alrededor 324.000.000 millones de comparaciones hechas por computadora. Para facilitar el procesamiento de datos, no se utiliza la vinculación de registros. En su lugar, el método propuesto se basa en clasificar los registros y compararlos únicamente con los registros consecutivos.

El *código Nombre* en el Gráfico 15 contiene los códigos de los nombres; el *código Paterno* y el *código Materno* son códigos para el primero y segundo apellidos. En el Gráfico 15 hemos reemplazado los nombres reales por códigos numéricos<sup>6</sup> – María podría tener el código 1076, González podría tener el código 1018, etc. En el gráfico, tenemos seis variables identificadoras con nombres y fechas de nacimiento. En una aplicación real, el lugar de nacimiento también debería incluirse.

**Gráfico15. Búsqueda de duplicados a través de la clasificación por identidades**

código Nombre	código Paterno	código Materno	AAAA	MM	DD	código combinado	igual
...	...	...	...	...	...	...	...
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 consideramos dos registros como duplicados. Empezamos la búsqueda de duplicados con la regla más estricta:

- Con las seis variables, se crea una combinación de código concatenando las 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 hemos encontrado un duplicado.
- Los dos registros que pertenecen al mismo duplicado son eliminados para tratamiento especial. En el caso del INEC, estos dos registros tienen números de cédula diferentes; otras

<sup>6</sup> Dado que el ejemplo se basa en datos reales, los nombres que aparecen son ficticios.

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 aceptamos un error o diferencia entre los códigos Paterno, Materno, año, mes y día de nacimiento podemos 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 a 100%. Si la similitud está por arriba de un límite, por ejemplo 80% y las otras variables identificadoras son iguales, entonces podemos tratar los registros como duplicados.
- Los duplicados definidos de acuerdo con una regla específica de similitud 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 residenciales incorrectos<sup>7</sup>**

Pensamos que los errores de cobertura son un problema que debería recibir mayor atención. Existe una larga tradición de trabajo con errores de muestreo y de no respuesta, y las tasas crecientes de no respuesta en las encuestas de Statistics Sweden han recibido mucha atención recientemente en los medios suecos. Sin embargo, los problemas de cobertura no han sido discutidos en los medios. El Registro de Población utilizado por Statistics Sweden tiene una cobertura excesiva y una cobertura insuficiente de entre 0.5%–1% para ambas.

El Registro de Empresas utilizado en las estadísticas económicas anuales tenía para el año 2004 una cobertura excesiva de alrededor de 8% y una cobertura insuficiente de alrededor de 28% en relación con el número de unidades legales y alrededor de 2% en relación con la rotación total (Wallgren y Wallgren (2014) pp.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 Sueco de Población, que genera serios problemas para las estadísticas por país de nacimiento. El problema aún no ha sido abordado.

En un país como Suecia, con muchos registros administrativos buenos disponibles, estos errores de cobertura se pueden reducir utilizando más fuentes y mejorando los métodos estadísticos. Sin embargo, en los países en desarrollo que empezaron recientemente a utilizar registros administrativos, se sugiere utilizar marcos duales para estimar los errores de cobertura, por ejemplo combinar registros con encuestas por muestreo basadas en marcos de área. El enfoque de doble marco para encuestas fue comentado por Ferraz (2015) y Carfagna y Carfagna (2010). También utilizaremos 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 de la aplicación. El alcance y los métodos utilizados para estas encuestas difieren de los métodos

---

<sup>7</sup> Esta sección se basa en Wallgren y Wallgren (2015b)

que discutimos aquí. En una encuesta posterior a la aplicación se hacen mediciones repetidas utilizando el mismo marco que en el censo. Los errores de mvalidación encontrados son utilizados para estimar errores por cobertura insuficiente en el censo. Un artículo reciente de da Silva et al. (2015) describe cómo se ha utilizado este método en el Censo Brasileño 2010.

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

Aquí se considera un caso diferente el cual 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 residenciales que no han sido actualizados hagan necesario combinar un registro estadístico con una muestra basada en marcos de 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.

Supongamos que queremos producir estadísticas regionales con el Registro de Población creado por un instituto nacional de estadística. El Gráfico 16 enlista 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 una población autóctona. Suponemos que estas variables pueden influir en el tamaño de los errores de cobertura para un municipio. Tal vez sea posible clasificar los hogares o las personas en diferentes categorías con diferente propensión a moverse a otro domicilio residencial. La edad, el nivel de educación y la actividad económica son variables que se pueden utilizar aquí.

**Gráfico 16. 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>	No está en el registro: <i>Idnr</i> = faltante	Número de identidad, <i>Idnr</i>
Domicilio residencial	Otras variables: igual que en el registro	Otras variables: igual que en el registro
A partir de la fecha de residencia	Es posible agregar variables estadísticas	y variables estadísticas para la encuesta
Distrito		
Municipio		
Categoría de municipio		
Sexo		
Edad		
Nivel de educación		
Actividad económica		
Categoría de persona o vivienda		

<sup>8</sup> En este artículo, preferimos utilizar el término *registro* en vez de *lista*.

Para una muestra basada en marcos de área, las mismas variables del registro son recolectadas por entrevistadores. Encontraremos personas en la muestra que no se encuentran en el registro o que tienen diferentes domicilios en el registro. Encontraremos cobertura insuficiente y cobertura excesiva a 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 basada en marcos de área y el registro: los porcentajes de personas con domicilio residencial 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 debe monitorearse con estas medidas.

El Gráfico 17 ilustra los diferentes tipos de errores de cobertura. A nivel nacional, la muestra basada en marcos de área proporciona una cobertura insuficiente estimada de 72.000 personas en el registro. La estimación de la muestra basada en marcos de área es de 80.000 persona que realmente viven en el distrito A, pero que están registradas como que viven en el distrito B. Estas 80.000 personas son una cobertura insuficiente de la subpoblación que vive en el distrito A y al mismo tiempo una cobertura excesiva de la subpoblación que vive en el distrito B.

Las personas que han salido del país es otra categoría de cobertura excesiva que no se puede detectar con los métodos discutidos aquí.

**Gráfico 17. Ejemplos de errores de cobertura, estimaciones por muestra por área de la población por distrito**

	Distrito según los datos del registro				No está en el registro	Todos	Cobertura insuficiente
	Distrito	A	B	C			
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
	Todos	<b>2 888 000</b>	<b>3 200 000</b>	<b>1 680 000</b>	72 000	<b>7 840 000</b>	330 000
Cobertura excesiva		18 000	90 000	150 000			

**Gráfico 18. Estimaciones no corregidas y corregidas**

Distrito	Estimaciones originales	Estimaciones Corregidas
A	2 888 000	3 079 000
B	3 200 000	3 193 000
C	1 680 000	1 568 000
Todos	7 768 000	7 840 000

En los Gráficos 17 y 18, se pueden comparar las estimaciones originales del registro de población en los distritos con las estimaciones de la muestra basada en marcos de área. En el registro, el distrito A está subestimado y los distritos B y C están sobreestimados. Podemos comparar las estimaciones del registro y las de la muestra basada en marcos de área para otras categorías de igual manera.

### Estimación con doble marco

Supogamos que la muestra por área nos 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 sus sean difíciles de utilizar para el caso que consideramos aquí (ver Ferraz (2015)):

- No creemos que la muestra basada en marcos de área se pueda diseñar de tal manera que se puedan obtener las estimaciones 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 exceso de cobertura. Los estimadores de doble marco sólo corregirán la cobertura insuficiente. Se detectará la cobertura excesiva del registro en la muestra basada en marcos de área cuando los entrevistadores conozcan a las personas que no han cambiado de domicilio en la muestra. En relación con las personas que han cambiado de domicilio en la muestra basada en marcos de área, los entrevistadores comprenderán que estos domicilios son incorrectos como domicilios residenciales de estas personas, pero será difícil determinar si el distrito o el municipio es correcto o no.

Nuestra conclusión es que los métodos de doble marco discutidos por Ferraz (2015) no son apropiados cuando tenemos tanto cobertura insuficiente como cobertura excesiva 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 las personas y establecimientos son ejemplos de dichas variables de clasificación. Por lo anterior, es necesario un enfoque de doble marco diferente 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 esta sección utiliza métodos similares al método de calibración utilizados para corregir la no respuesta en encuestas por muestreo. La muestra basada en marcos de área discutida en la sección anterior se puede utilizar de dos formas:

- Para proporcionar estimaciones de los errores de cobertura y los errores en las variables de clasificación básica en el registro, de acuerdo a lo descrito en la sección anterior sobre la estimación de doble marco.
- Para proporcionar estimaciones que puedan ser utilizadas como condiciones de calibración para ajustar o corregir errores de cobertura.

El Gráfico 19 muestra la estructura del Registro de Población. Las variables de las columnas (2)–(6) se pueden utilizar para definir las condiciones de calibración si queremos ajustar por errores de cobertura. Si elegimos utilizar únicamente *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).

**Gráfico 19. 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 sumamos la columna (7) por distrito, obtendremos las estimaciones originales (sin corregir) del Gráfico 20. Si sumamos la columna (11) por distrito, obtendremos las estimaciones corregidas del Gráfico 20:

**Gráfico 20. Estimaciones no corregidas y corregidas**

Distrito	Estimaciones originales	Estimaciones corregidas
A	2 888 000	3 079 000
B	3 200 000	3 193 000
C	1 680 000	1 568 000
Todos	7 768 000	7 840 000

En este caso, es fácil calcular los pesos calibrados. Los pesos  $w_i$  para los registros con Distrito A en el registro serán  $3079/2888 = 1.066$  y los pesos para los otros distritos se calculan de manera similar. Con las columnas (8) – (10) los mismos pesos se pueden derivar con álgebra matricial.

Los datos de la muestra basada en marcos de área deben analizarse 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 ser utilizado para ajustar o corregir las estimaciones por errores de cobertura. Este trabajo metodológico será muy similar al trabajo hecho cuando se ajustan los datos por no respuesta en las encuestas por muestreo.

Si existen más registros estadísticos con datos sobre, por ejemplo, educación, empleo o ingresos, estos registros deben utilizar la población definida por el Registro de Población como sus poblaciones de registro. Los pesos calibrados  $w_i$  deben utilizarse para la estimación de estos registros también. Estos registros 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 de muestra 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 son corregidas tanto por no 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, combinamos la información de la muestra por zona con la información de la muestra basada en registros con el fin de obtener estimaciones corregidas o ajustadas por errores de cobertura. Hemos vinculado un marco de área con un marco de lista con 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 consideramos estadísticas basadas en registros y utilizamos la muestra por áreas para corregir las estimaciones basadas en registros.

### ***Discusión***

Muchos países tienen Registros de Población, Registros de Empresas y Registros de Explotación Agrícola que tienen tanto problemas de cobertura insuficiente como de cobertura excesiva. La información básica de registros, como domicilio residencial o actividad económica, puede estar desactualizada o incorrecta. Estos factores de calidad deben ser monitoreados por muestreo de áreas. La calidad de las estimaciones basadas en registros también se puede mejorar calibrando los pesos. 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 deben mejorarse para reducir gradualmente los problemas de cobertura.

Hoy en día publicamos estimaciones de encuestas por muestreo con tasas de no respuesta de 40% o incluso mayores. Podemos hacerlo porque confiamos en nuestros métodos para ajustar estimaciones por no respuesta con la información de los registros que usamos como marcos. ¿Podría-

## Conclusiones

mos también ajustar las estimaciones basadas en registros cuando tenemos un 40% de cobertura insuficiente? Para nosotros, esta idea es nueva y alarmante. Como estadísticos de Suecia, estamos acostumbrados a una situación donde la cobertura insuficiente es cuando mucho de 0.5%–2%. Esta cobertura insuficiente es considerada tan pequeña que es despreciable. En primer lugar, queremos aclarar que ahora comprendemos que los errores de cobertura insuficiente pueden ser muy selectivos y que la cobertura excesiva o insuficiente entre 0.5%–2% *no* se debe despreciar ya que puede generar errores graves. También recordamos que cuando estudiamos estadística en nuestra juventud de estudiantes, nos enseñaron que si la tasa de no respuesta era mayor a 5%, ¡la encuesta por muestreo era inútil!

Nuestra conclusión es que las actitudes hacia los errores en las encuestas cambian con el tiempo. Si tenemos confianza en nuestros métodos para ajustar las estimaciones por no respuesta o por errores de cobertura, podemos hacer correcciones y publicar. Corregir las estimaciones por no respuesta y corregir por errores de cobertura con el método que proponemos aquí requiere de métodos muy similares. Por lo tanto, nuestras recomendaciones son:

- 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,
- corregir las estimaciones y publicar si confían en su método de corrección.

## 6 Conclusiones

Regresamos a las conclusiones de la Sección 2.5 y agregamos algunos comentarios:

*Si se usan registros administrativos para la producción de estadísticas, se pueden ahorrar costos y mejorar la calidad. Los censos pueden ser remplazados por encuestas por registros, y se pueden desarrollar nuevos tipos de estadísticas regionales y longitudinales. El sistema nacional puede cambiarse a un sistema estadístico 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 basadas en marcos de área.*

Si esta visión se convierte en realidad, se requerirá una cooperación permanente entre secretarías, institutos nacionales y autoridades. Tanto la alta dirección como el personal que trabaja con los registros administrativos deben participar en esta cooperación. Además, el trabajo dentro del instituto nacional de estadística debe organizarse de una nueva manera. El modelo de proceso de producción presentado en la Sección 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 trabajo y mejorar la consistencia. Los gerentes del instituto de estadística deben participar en esta coordinación.

Hay un comentario en la página 22 sobre cómo se puede mejorar la calidad:

*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, dando la oportunidad de encontrar y reducir errores.*

Hemos encontrado que este principio que parece sencillo es muy difícil de implementar. Las personas que trabajan con diferentes encuestas tienen diferentes áreas de interés y competencias. Además, más de un gerente estará involucrado; cada uno con su propio presupuesto. La encuesta mensual de negocios por muestreo y el registro anual con datos de empresas en Statistics Sweden pueden ilustrar esto:

En noviembre de 2013, se definió la población marco para la encuesta por muestreo de enero de 2014. No se incluyeron las pequeñas empresas en este marco. Se llevaron a cabo encuestas por muestreo similares para febrero-diciembre de 2014.

Durante septiembre de 2015, una unidad de registro en Statistics Sweden recibirá declaraciones de ingresos para enero-diciembre de 2014 de todas las empresas. Entonces se pueden comparar las poblaciones y variables de las 12 encuestas mensuales por muestreo con el registro anual. Los errores de marco en las encuestas por muestreo pueden ser graves, y 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 comparando 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 es utilizado a su máximo potencial.

---

<sup>9</sup> En Wallgren y Wallgren (2014) los errores de cobertura que se pueden detectar y corregir se discuten en las páginas 130-140.



Estas experiencias explican el por qué recomendamos la organización de una unidad central de registro. Esta unidad debe tener un panorama general del sistema de producción y medir y reducir los errores e inconsistencias. Los institutos líderes de estadística en países como Estados Unidos y Canadá 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 en otros países, lo cual tal vez explica por qué los metodólogos y los gerentes en muchos países aún no han entendido el enfoque de sistemas.

## Referencias

- Carfagna E. and Carfagna A. (2010) Alternative sampling frames and administrative data; which is the best data source for agricultural statistics? In Benedetti, Bee, Espa, Piersimoni (Editors), *Agricultural Survey Methods*, Wiley & Sons, Ltd, Chichester, UK.
- Deville, J. and Särndal, C-E. (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376–382.
- FAO (2010) Estrategia Global para el Mejoramiento de las Estadísticas Agropecuarias y Rurales. *Informe número 56719-GLB. Banco Mundial 2010*.
- Ferraz, C. (2015) Linking Area and List Frames in Agricultural Surveys. Paper written for FAO.
- Laitila, T., Wallgren, A. and Wallgren B. (2012) *Quality Assessment of Administrative Data – Data Source Quality*. Part two of third deliverable of work package 4 of the BLUE Enterprise and Trade Statistics project. Can be obtained from the authors: [ba.statistik@telia.com](mailto:ba.statistik@telia.com)
- Silva A., Soares M., Carneiro D. (2015): Assessing coverage of the 2010 Brazilian Census. *Statistical Journal of the IAOS*, 31, 215-225.
- Statistics Canada (2009) *Statistics Canada Quality Guidelines*, 5th edition.
- Stender P., Thorsen T., Andersen H. (2015): Micro data integration for Labour Market Account. *Statistical Journal of the IAOS*, 31, 249-256.
- Wallgren A., Wallgren B. (2012) *Estadísticas basadas en registros. Aprovechamiento estadístico de datos administrativos*. INEGI Mexico.
- Wallgren A., Wallgren B. (2014) *Register-based Statistics – Statistical Methods for Administrative Data*. Second edition, John Wiley & Sons, Ltd, Chichester, UK.
- Wallgren A., Wallgren B. (2015a): Quality assessment in production systems with registers and sample surveys. *Statistical Journal of the IAOS*, 31, 241-247.
- Wallgren A., Wallgren B. (2015b): Frames and populations in a register-based statistical system. *Paper presented at ISI 2015*.