

El uso de nuevas tecnologías para la producción de estadísticas públicas. Un camino recorrido y por recorrer

María Teresa Arriola, Emilio Platzer y María Eugenia Lago

Introducción

La DGEyC coordina y dirige los servicios que conforman el Sistema Estadístico de la Ciudad (SEC), como lo establece la Ordenanza 35.386/79. Sus proyectos estadísticos son de naturaleza censal, encuestas por muestreo y provenientes de registros administrativos. Desde 2007, el organismo se propuso alcanzar estándares de calidad internacional. En tal sentido, ese mismo año certificó todos sus procesos de trabajo (ISO) y, paralelamente, inició un proceso de incorporación de nuevas tecnologías para aumentar la calidad de los datos (relevancia, exactitud, exhaustividad, oportunidad, accesibilidad, comparabilidad y coherencia).

Dentro de la incorporación de nuevas tecnologías, se destaca el desarrollo del Sistema Integrado de Encuestas a Hogares (SIEH) de la Ciudad de Buenos Aires. Se desarrolló un software específico, se modernizaron los procesos de trabajo, se comenzaron a utilizar dispositivos móviles para la captación del dato. Todo esto permitió mejorar la calidad de los datos, acelerar los tiempos de producción de resultados y aumentar la capacidad de ejecutar mayor cantidad de operativos.

María Teresa Arriola es Master en Coaching y Cambio Organizacional y Licenciada en Ciencias de la Computación. Se desempeña como Directora de Estadísticas Sociodemográficas de la Dirección General de Estadística y Censos de la Ciudad Autónoma de Buenos Aires (DGEyC-CABA).

Emilio Platzer es Licenciado en Ciencias de la Computación. Se desempeña como Jefe de Departamento Proyectos Especiales de la Dirección General de Estadística y Censos de la Ciudad Autónoma de Buenos Aires (DGEyC-CABA).

María Eugenia Lago es Magíster en Demografía Social, Profesora y Licenciada en Sociología. Se desempeña como Subdirectora General de Estadísticas Sociodemográficas de la Dirección General de Estadística y Censos de la Ciudad Autónoma de Buenos Aires (DGEyC-CABA).

De la necesidad del dispositivo móvil al desarrollo de un sistema integrado de encuestas a hogares

Durante el año 2010, maduró dentro de la DGEyC la idea de utilizar dispositivos móviles para la captación de los datos de las encuestas. A principios de 2011, se definieron los objetivos que dieron inicio al proyecto de desarrollo del SIEH.

El uso de dispositivos móviles debía ayudar a los encuestadores a evitar los principales errores (la omisión de bloques de preguntas, errores en el flujo y filtros del cuestionario), acelerar el ingreso del dato al sistema informático (por la eliminación los errores de transcripción) y realizar la retroalimentación inmediata a los encuestadores (mejorando la calidad).

Se decidió empezar a implementar el uso de dispositivos móviles con la Encuesta Anual de Hogares, que es una encuesta multipropósito que realiza la DGEyC desde el año 2002 y cuyo tamaño muestral está compuesto por 9.000 viviendas.

En 2011, la producción de la EAH estaba parcialmente informatizada utilizando distintas herramientas informáticas cuya interconexión era realizada por distintos procesos manuales. Es por ello que, si bien era posible acotar el proyecto a la utilización de un cuestionario digital, se decidió avanzar en la informatización de otras etapas de la producción de las encuestas, principalmente el procesamiento.

El desarrollo del sistema informático debía no solo incluir la información de las encuestas relevadas sino también administrar el trabajo de campo y el procesamiento.

Finalmente, se decidió encarar un proyecto que desarrollara un Sistema Integrado de Encuestas a Hogares, basado en un sistema informático que incluyera el ciclo completo de la vida de la encuesta (desde la selección de la muestra hasta la publicación de todos sus resultados).

Se plantearon objetivos a corto y a largo plazo:

- A largo plazo: la inclusión de todo el ciclo de vida de todas las encuestas a hogares de la DGEyC.
- A corto plazo: el desarrollo de un sistema para la Encuesta Anual de Hogares: que permitiera

controlar la trazabilidad del dato; en donde pudiera coexistir el relevamiento en dispositivos móviles con el relevamiento en papel; que administrara el trabajo de campo; y que incluyera un módulo de edición de encuestas para continuar con las tareas de recepción.¹

A fines de 2011, se realizó una prueba en campo de levantamiento de datos con dispositivo móvil; en el primer semestre de 2012, se llevó a cabo una prueba piloto; y en el segundo semestre se incorporó definitivamente en la EAH el dispositivo móvil. Eso implicó el consecuente desarrollo del sistema integrado cuyos módulos se fueron agregando secuencialmente en el tiempo dando lugar a la concreción del objetivo a largo plazo. En los años subsiguientes, se fueron incorporando distintas etapas de la producción estadística y se produjeron distinto tipo de encuestas (Gráfico 1).

¹ Véase en el Anexo roles y funciones de la estructura operativa de campo.

Gráfico 1
Ciclo de vida de un proyecto estadístico muestral



¿Qué es el Sistema Integrado de Encuestas a Hogares de la Ciudad de Buenos Aires?

- Diseño conceptual de marcos compartidos: aspectos básicos comunes entre encuestas y aspectos diferentes sobre la base de un criterio armonizado.
- Marco maestro de muestreo unificado con estrategias específicas (con actualización cartográfica permanente).
- Operativos de campo coordinados en una estructura única (compartiendo pool de coordinadores, supervisores y capacitadores).
- Procesamiento de la información sobre la base de criterios unificados que contemplan la definición de todos los aspectos (diseño conceptual, muestra, operativo de campo).
- Sistema informático que incluye la información de todas las áreas desde la muestra, registrando las definiciones conceptuales, la logística de campo, el control de consistencias durante el relevamiento y el procesamiento y control de la información hasta la producción de tabulados y base a usuario.

La muestra

Como lo indican Feres y Medina (2001), uno de los aspectos más importantes de un sistema integrado es contar con una muestra maestra. En 2013, sobre de la base de datos del Censo Nacional de Población, Hogares y Viviendas 2010, se planteó una estrategia de un diseño muestral para satisfacer una nueva encuesta –la Encuesta Trimestral de Ocupación e Ingresos (ETOI)– y mantener los diferentes dominios de estimación de la EAH y las encuestas especiales demandadas por otros organismos del Gobierno de la Ciudad de Buenos Aires y del orden Nacional.

Con ese fin, se definieron y estratificaron² nuevas Unidades Primarias de Muestreo (UP) y se seleccionaron las UP para satisfacer la demanda de todos los proyectos estadísticos señalados.

Paralelamente, se definió una estrategia de Listado para mantener actualizadas las viviendas de las UP seleccionadas.

El Diseño conceptual sobre la base de un sistema basado en metadatos

Se define como “datos” del sistema a las viviendas, hogares y personas encuestados (por ejemplo, a las características de las viviendas, nombres, edades, ingresos, situación laboral y nivel educativo de las personas).

Los metadatos del sistema están conformados por los texto de las preguntas, los nombres de las variables donde se almacenan las respuestas, el tipo de datos esperado para cada uno, los códigos y textos de las opciones de las preguntas cerradas, los rangos posibles en las variables numéricas abiertas y los destinos de los saltos de las opciones de preguntas que determinan pases.

El sistema informático almacena los datos y los metadatos de la encuesta en la misma base de datos. Eso permite utilizar los metadatos para desplegar los cuestionarios, hacer controles de información válida, correr pautas de consistencias, etcétera.

Los metadatos del SIEH (Figura 1) conforman una estructura que permite almacenar las definiciones lógicas de las unidades de análisis (viviendas, hogares, población), las dimensiones, las variables y los posibles valores que asumen. En este sentido, al diseñar el cuestionario, se diseña la base de datos. La principal ventaja de este tipo de programación es que no requiere modificar los programas

² El diseño responde a un modelo bietápico: selección de Unidades Primarias de Muestreo (UP) (radios censales); estratificación de las UP en cinco estratos, cada uno equivalente a 5 quintiles de ingreso; y selección de las UP con probabilidad proporcional al tamaño (método de Sampford).

(ni hacer programas nuevos) cuando se decide realizar otra encuesta de características similares a las ya realizadas. Así, el primer requisito de un nuevo proyecto estadístico es la definición de los metadatos en el sistema, que, una vez controlados y revisados, dan inicio al operativo sin necesidad de cambios en la programación. Solo es necesario realizar cambios en el sistema informático cuando se desea agregar un comportamiento distinto, nuevos tipos de preguntas o de filtros, etcétera.

Datos

El dato es el producto del registro de algún tipo de acción manifiesta de observación. El dato son los valores que asume una variable, las dimensiones o características al describir una unidad de análisis/observación. En el caso del SIEH, los datos describen las viviendas, los hogares y sus miembros (tipo de vivienda, cantidad de ambientes utilizados por el hogar, edad de las personas que viven en el hogar, etcétera).

Metadatos

Los metadatos son los datos que describen las características de los datos. Son los datos sobre los datos. En el caso del SIEH, los metadatos describen a los cuestionarios, preguntas, variables, etcétera.

Se almacena como metadatos la información sobre:

- 1.- los cuestionarios, el listado de preguntas (código, textos, tipo de respuestas, saltos), las opciones de las preguntas (número, texto y destino del pase si lo tuviera), los filtros (condiciones, textos, pases y secciones);
- 2.- la estructura de la base de datos, nombres de tablas donde se almacenan las unidades de análisis (vivienda, hogar, personas, etc.), nombres y tipos de variables donde se almacenan las respuestas de los cuestionarios;

Figura 1

Sistema Integrado de Encuestas a Hogares (SIEH): Módulo de metadatos del cuestionario

Grilla: preguntas , registros: 12						
pre	texto	for	mat	bio	aclaramiento	orden
		Δ ⊕ + Δ ⊕ Δ ⊕			Δ ⊕	Δ ⊕
F		I1		T		
B T1	¿La semana pasada trabajó _____ por lo menos una hora?	I1		T		18
B T2	En esa semana, ¿hizo alguna changa, fabricó en su casa algo para vender, ayudó a un familiar o amigo en su negocio?	I1		T		20
B T3	¿La semana pasada...	I1		T	G-S. Primero lea todas las opciones y luego marque la respuesta	30
Grilla: opciones , registros: 5						
opc	texto	aclaramiento		destino		
	Δ ⊕	Δ ⊕				
B 1	no deseaba, no quería trabajar?			I20		
B 2	no podía trabajar por razones personales?	(estudio, cuidado del hogar, etc.)		T9		
B 3	no tuvo pedidos/ clientes?			T9		
B 4	no tenía trabajo y quería trabajar?			T9		
B 5	tenía un trabajo/ negocio al que no concurrí?					
B T4	¿No concurrí a su trabajo por...	I1		T	G-S	40
B T5	¿Le siguen pagando durante la suspensión?	I1		T		50

- 3.– las fórmulas de las variables calculadas (por ejemplo, cantidad de hijos menores en el hogar, condición de actividad, años de escolaridad, etc.);
- 4.– las pautas de consistencia;
- 5.– la estructura de los cuadros de resultados y tabulados de control;
- 6.– la selección de columnas que formarán la base usuario y/o bases internas de trabajo;
- 7.– el flujo del trabajo de campo, la forma de organización del campo por semanas o lotes, condiciones o cuotas de recuperación y/o supervisión de encuestas.

Sobre las características del cuestionario digital

En el año 2011, eran pocas las experiencias de utilización de dispositivos móviles para el relevamiento de encuestas. La mayoría utilizaba notebooks, otras utilizaban dispositivos que habían dejado de fabricarse o que eran hardware específicos cuya continuidad en el mercado los proveedores no podían estimar. Las notebooks (laptops, PC portátiles y las llamadas en ese momento netbooks) tenían dos problemas importantes: el peso y la autonomía de la batería. Las más livianas de un kilo o más tenían poca capacidad y poca autonomía de batería. Las que tenían más autonomía (hasta 6 horas) y capacidad eran mucho más pesadas, de dos kilos y medio o más.

La DGEyC decidió utilizar tabletas³ después de haber evaluado las características físicas y de haber probado 5 modelos distintos, utilizando un prototipo de encuesta programado para esta tarea.

³ En abril de 2010, Apple había desarrollado el iPad, y, a principios de 2011, se conseguían en Argentina tabletas de distintos proveedores. Se disponía entonces de un dispositivo de menos de 800 gramos de peso con una autonomía de batería de más de 8 horas que, gracias a su facilidad y velocidad para pasar del modo encendido al modo de reposo, hiciera mayor la cantidad de horas reales de uso. Esas características lo hacían ideal para el trabajo habitual de nuestros encuestadores, que muchas veces realizan las encuestas parados en la calle (cuando el encuestado contesta a través de una reja) sin posibilidades de apoyar el dispositivo ni de enchufarlo.

Las primeras tabletas tenían una pantalla de 10" y un peso aproximado de 700 gramos; luego aparecieron tabletas de entre 7 y 8" con un peso aproximado de 350 gramos, el mismo poder de cómputo y autonomía de batería. Actualmente, la DGEyC cuenta con tabletas de ambos tamaños.

La coexistencia entre el papel y el dispositivo móvil

Al inicio del proyecto, el equipo de campo tenía la inquietud de cómo proceder en el caso de que no se pudiera iniciar o continuar la entrevista con el dispositivo móvil (falla del equipo, agotamiento de la batería, cuestiones de seguridad). Se decidió la coexistencia del relevamiento en papel y en dispositivo móvil. Los encuestadores llevarían a campo un dispositivo móvil y formularios en papel y estarían capacitados en el uso de ambos soportes de relevamiento. Se sostuvo que, ante cualquier dificultad, utilizaran el papel (ya fuera para comenzar una encuesta nueva o para continuar una empezada). El fundamento de la decisión tomada fue que el objetivo era la captación de los datos y no la utilización de un soporte u otro. En este sentido, conceptualmente, la tecnología debe adaptarse a mejorar la tarea y no la tarea adaptarse a las limitaciones tecnológicas.

Para unificar criterios de validación entre las encuestas relevadas en papel y las del dispositivo móvil, los cuestionarios que se hubieran realizado en papel se ingresan en el sistema informático en el momento en que el encuestador llega y entrega su carga de trabajo; de esta manera se aprovechan los controles automáticos. El recepcionista realiza su tarea de control de las encuestas recibidas a medida que se van ingresando de modo de aprovechar los controles automáticos del sistema y la presencia del encuestador. Así las tareas de ambos no son muy diferentes en papel y en dispositivo móvil. Cuando el ingresador terminó de incorporar los cuestionarios de la primera vivienda recibida, el recepcionista puede empezar su tarea.

El diseño del programa para el dispositivo móvil

Se utilizó como guía para el diseño del software la teoría sobre Usabilidad del Software desarrollada por Jakob Nielsen a principios de la década de 2000 y los siguientes criterios generales acordados en la DGEyC: que el dispositivo móvil fuera en la funcionalidad lo más parecido al papel; que ayudara al encuestador a realizar los saltos de preguntas en la administración de pases y filtros del cuestionario; que la redacción de preguntas y opciones de respuesta fuera idéntica; que no agregara pausas o demoras mayores a las que hay en papel (al cambiar de hoja o abrir otro cuestionario); y que en cualquier momento de la entrevista, o aun después de finalizada, permitiera retroceder, corregir o borrar (hasta el momento de entregar los cuestionarios al recepcionista o de descargar/transmitir las encuestas del dispositivo móvil).

Siguiendo los principios de usabilidad, las opiniones del personal con experiencia de la Subdirección General de Estadísticas Sociodemográficas, el consejo de los expertos y los antecedentes (otros desarrollos para PC y dispositivos móviles), se diseñaron alternativas factibles para una serie de características del software. Luego se evaluaron las alternativas haciendo pruebas con usuarios de la subdirección.

Las principales alternativas evaluadas fueron:

1. *Una pregunta por pantalla versus un conjunto de preguntas.* Hacer una sola pregunta era la única alternativa en pantallas pequeñas (del tamaño de un celular chico) y era una alternativa muy usada en los desarrollos anteriores. Su ventaja era la facilidad de programación y la focalización del encuestador en la pregunta a formular. No obstante, la alternativa elegida fue la visualización de varias preguntas porque se observó que facilitaba la ubicación del encuestador al mantener una visión del contexto de la

pregunta, introducía menos errores, daba más seguridad y permitía un retroceso ágil a la hora de corregir o consultar una respuesta previamente marcada. La alternativa elegida conserva las características del cuestionario en papel.

2. *Visualización parcial o completa del cuestionario.*

Se decidió visualizar el cuestionario completo en lugar de visualizarlo por páginas, bloques, capítulos o conjuntos de preguntas. El cuestionario se despliega completo en una “única página continua” con desplazamiento vertical. La decisión se tomó por ser esta la opción más ágil.

3. *Mostrar versus ocultar las preguntas saltadas en los pases.*

Se eligió mostrar las preguntas saltadas y las por contestar con un color diferente para que el encuestador distinguiera la pregunta a realizar de las que debía omitir. Esto mejoró la visualización de contexto y la detección de errores en opciones de preguntas que generan un salto (porque el encuestador al ver el bloque saltado tiene la oportunidad de darse cuenta de que marcó mal). Esta decisión tuvo dos beneficios adicionales: la posibilidad de imprimir el cuestionario completo; y facilitar la capacitación de los encuestadores gracias a la posibilidad de mostrar todos los flujos posibles del cuestionario digital como lo permite el papel.

4. *Borrado automático versus borrado manual.*

En el desarrollo de la encuesta, el encuestador podría tener que corregir una pregunta que fue realizada más arriba en el cuestionario. Si la pregunta que debe modificarse implica un cambio en el flujo del cuestionario, los datos de las preguntas saltadas deben borrarse. El riesgo de borrar automáticamente es que esa corrección puede ser errónea y, entonces, cuando el encuestador rectifica el cambio encontraría que los datos fueron eliminados y debería repreguntar. El borrado manual permite al encuestador detectar su error al ver qué respuestas debe eliminar.

5. *Posibilidad de marcar fuera de flujo.* Se decidió permitir marcar opciones de respuesta en las preguntas que por el flujo de pases debían ser omitidas, salteadas. Esta decisión se tomó pensando en atenuar el impacto de posibles errores de programación de pases o de carga de metadatos que impidan consignar respuestas a preguntas que el encuestador tiene la certeza de que debe formular (dada su experiencia o la capacitación recibida). La alternativa de impedir marcar implicaría perder el dato u obligarlo a consignarlo en un lugar que no es específico (por ejemplo, el campo observaciones). Por otro lado, la desventaja del criterio adoptado es que permite tomar al encuestador una decisión que puede resultar errónea. Luego de un año de probar la primera alternativa, se decidió mantener esta característica y permitir siempre consignar respuestas aun en preguntas salteadas porque, si bien los errores fueron casi inexistentes (tanto de programación como de criterio de los encuestadores), la ventaja de permitir supera a la de impedir. Para minimizar los errores, el programa mantiene y calcula el flujo normal y marca con un color especial todas las preguntas que se contestan “fuera del flujo normal”.

6. *Flexibilidad en el orden de administración de cuestionarios.* Se decidió por ser flexible (al igual que el papel) en la administración de cuestionarios. Por ejemplo, se puede suspender la administración de un formulario a un miembro del hogar y comenzar con el cuestionario de otro miembro del hogar, para luego continuar con el suspendido en el punto en que fue dejado. La flexibilidad de uso del dispositivo móvil es independiente de la metodología de captación que indica que el encuestador debe realizar las preguntas en el orden en que figuran, finalizar cada cuestionario antes de empezar el siguiente o terminar el renglón de la persona (en la matriz de componentes del hogar) antes de empezar el siguiente. La implementación de los

múltiples flujos posibilita interrumpir el flujo de indagación en distintos momentos. El programa controla la completitud de los cuestionarios, marca con color los incompletos de modo que, al final de la tarea, deba haber completado las partes faltantes y borrado las sobrantes para que la encuesta no quede marcada como inconsistente o incompleta.

Las consistencias

Las pautas de consistencia son los controles que se aplican a cada una de las encuestas realizadas. Comparan dos o más datos que individualmente pueden ser correctos, pero que en conjunto no lo son o es muy improbable que lo sean. Por ejemplo, una pauta es: la edad de una persona no puede ser mayor que la edad de la madre.

Tradicionalmente, las pautas de consistencias se diseñaban para ser utilizadas dentro de la oficina por el equipo de procesamiento; a veces se incluían en el sistema de ingreso (en especial, cuando se usaba la figura del analista-ingresador). El recepcionista también tenía un manual de pautas de consistencia para controlar manualmente los cuestionarios recibidos. Las pautas de consistencia del recepcionista eran un subconjunto de la lista total porque, al aplicarlas manualmente, se hacía imposible administrarlas en su totalidad. Al diseñar una encuesta, se evaluaba qué pautas debían administrarse en la recepción según su importancia o probabilidad de ocurrencia.

A partir del dispositivo móvil, las consistencias se pudieron aplicar durante el desarrollo de la entrevista. Así se garantizaría que toda la encuesta fuera consistente desde el momento del relevamiento.

Si bien existía un vasto listado de pautas de consistencias al iniciar al desarrollo del SIEH, su incorporación al dispositivo móvil implicó nuevas definiciones. Hubo que evaluar la pertinencia

de la inclusión de cada una para no entorpecer la entrevista. También se tuvo en cuenta que el encuestado podía estar mirando la pantalla del dispositivo móvil, por lo cual se tuvo especial cuidado en la elección de los colores y los textos.

Para automatizar las pautas de consistencias, se escribieron como una expresión lógica entre variables con una precondition y una postcondición (en el ejemplo, la precondition sería si la madre está en el hogar, la postcondición será “la edad debe ser menor que la edad de la madre”) (Figura 2). Las pautas de consistencias se clasificaron:

- según el momento en que se muestran: en *relevamiento*: son mostradas en el dispositivo móvil (estas son la gran mayoría); en la *recepción* (el recepcionista le pregunta al encuestador y anota en el sistema lo que el encuestador explique al respecto); y en el *procesamiento*.

- según la posibilidad de ocurrencia: son *errores* cuando lo consignado no es considerado posible; son *advertencias* cuando lo consignado es muy poco probable.

El trabajo de revisión y definición de pautas de consistencias se realizó dentro del Sistema Informático con un equipo multidisciplinario (sectorialistas, diseño conceptual, procesamiento, campo, etc.).

La capacitación de encuestas con dispositivo móvil

La capacitación de los recursos humanos de la estructura operativa tuvo y tiene tres destinatarios: encuestadores, recepcionistas y coordinación de campo, entendiendo que cada perfil usa funcionalidades diferentes y específicas inherentes a su rol. No obstante, desde el inicio, se consideró fundamental que el conocimiento fuera acumulativo a medida que el rol fuera de mayor jerarquía que el anterior.

Figura 2
Sistema Integrado de Encuestas a Hogares (SIEH): Módulo de edición de consistencias

Grilla: consistencias, registros: 4								
A)	precondicion	rel	postcondicion	activa	explicacion	falsos positivos	momento	valida
B)	Δ e	Δ e	Δ e	Δ e	Δ e	Δ e	Δ e	Δ e
F)								
B)C	TOTAL_H=1 or v1=1 =>		H3=V4	Sí	Si es un solo hogar, la cantidad de habitaciones de la vivienda (V4) debe ser igual a la cantidad de habitaciones de uso exclusivo (H3)	no	Relevamiento 1	Sí
B)C	P4 =1	=>	EDAD>=18	Sí	El jefe preferentemente debe tener 18 años o más	Sí	Relevamiento 1	Sí
B)C	P4=2	=>	P5=1 or P5=2	Sí	Si es cónyuge, debe ser unido/a o casado/a	no	Relevamiento 1	Sí
B)C	T38=1	=>	T44=3	Sí	//*Si trabaja en el estado (t38=1), no puede ser patrón ni cuentapropia ni trabajador familiar (t44 a t48)	no	Procesamiento	Sí

Inicialmente, se armó material didáctico específico y se asignó un espacio propio dentro del plan de clases a la explicación del uso del cuestionario electrónico. Durante los primeros dos años, se capacitó a los encuestadores: en una primera etapa, en aspectos conceptuales utilizando el cuestionario en papel; en una segunda etapa, en aspectos funcionales del dispositivo móvil, para finalizar con una simulación de entrevista.

Posteriormente, y con la naturalización del uso de dispositivos táctiles, se observó que era suficiente con la inclusión, dentro de la capacitación conceptual, de la información correspondiente a cada módulo del sistema.

Dado que era muy beneficioso hacer significativo el sistema informático a partir de su aplicación a las tareas definidas conceptualmente, una estrategia didáctica aún vigente es el *role playing* que se realiza sobre el dispositivo móvil ya que hace emerger inquietudes operativas y conceptuales en un espacio lúdico y distendido.

La organización del trabajo de campo dentro del SIEH

El estado de las encuestas

Las encuestas en papel existen físicamente; cada encuesta está en un único lugar posible (en campo, en la recepción o en supervisión para volver a campo o en ingreso). Una vez ingresados los datos, el papel se archiva y pasa a existir de manera digital dentro del sistema.

Cuando se utilizan dispositivos móviles, la información de la encuesta puede estar simultáneamente en el dispositivo móvil del encuestador y en la base de datos centralizada del servidor. Es por eso que el sistema debe mantener una “marca lógica” indicando que el “poseedor” de la encuesta es el encuestador. El “poseedor” es el único autorizado a modificar los datos de la encuesta.

Cuando la encuesta es descargada del dispositivo móvil, su estado puede ser variado: completa, incompleta, inconsistente, no respuesta, rechazo, hogar ausente, domicilio inexistente o no visitado aún; en función del estado y de la organización definida en el flujograma de la organización de campo de la encuesta, la encuesta se vuelve a cargar para el mismo encuestador, es mandada a recuperar o a supervisar o pasa al sector de procesamiento (Gráfico 2).

Pertenece a los metadatos de la organización de campo la información referente a:

1. los roles que participan en la encuesta: encuestador, recuperador, supervisor, supervisor de recuperador, recepcionista, analista de procesamiento, etcétera;
2. el orden y los tiempos que tiene cada rol para hacer su tarea;
3. los códigos de respuesta/no respuesta y las tareas que se deben hacer a continuación basadas en ellos;
4. los permisos que tiene cada rol para editar la encuesta según el estado en el que esté;
5. el tipo de organización de la muestra: por períodos de tiempo (semanal en caso de la EAH) o por lotes (para algunas encuestas puntuales).

Con estos metadatos, el sistema, además de saber qué dispositivos tiene que cargar y descargar, puede hacer el control agregado de campo (por período o por semana) para ir monitoreando la tasa de respuesta, el nivel de cobertura, el ajuste con los cronogramas, etcétera.

El sistema contiene el listado de todas las viviendas que deben visitarse. La EAH estaba organizada en 3 etapas de un mes cada una. En el sistema, las viviendas se agrupan en lotes de trabajo que son

Gráfico 2
Estados de las encuestas

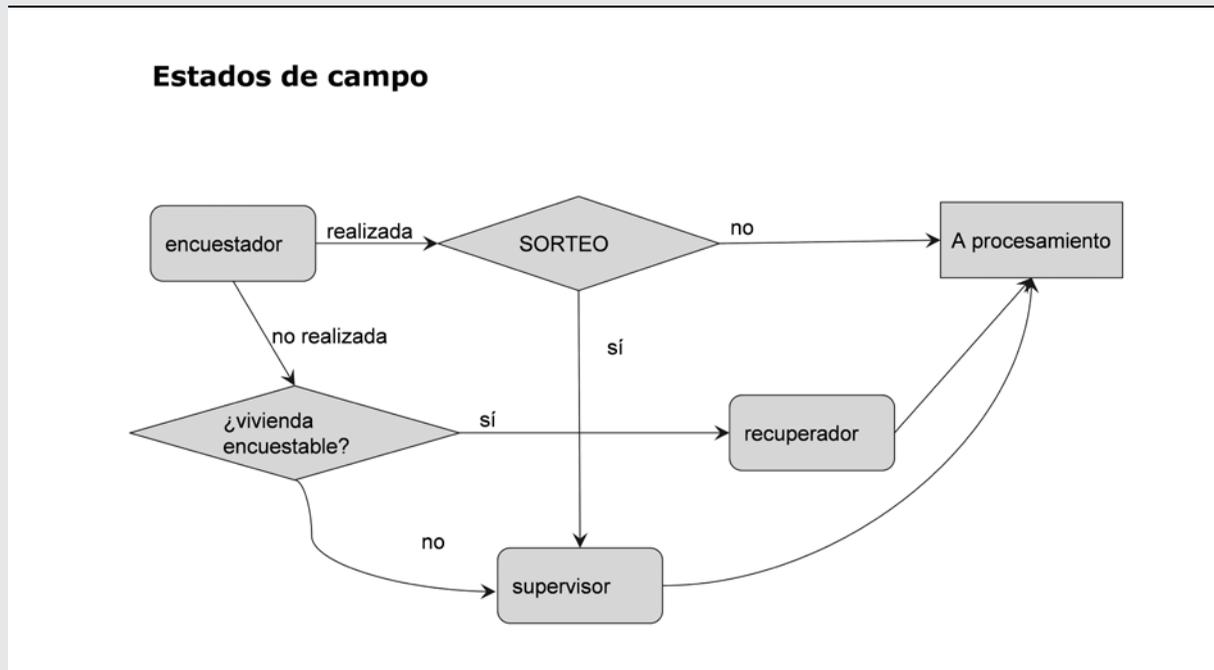


Figura 3
Sistema Integrado de Encuestas a Hogares (SIEH): Tabla de estados

est	ope	nombre	editar encuesta
18	etoi153	En reserva	nunca
20	etoi153	Disponible para ser asignada por el recepcionista.	nunca
22	etoi153	Asignada a encuestador.	nunca
23	etoi153	Cargada en el (DM); en campo para encuestador.	descargando
24	etoi153	Cargada en el (papel); en campo para encuestador.	nunca
25	etoi153	Descargada vacía para ingreso o para volver a cargar	nunca
26	etoi153	En ingreso (de encuestador)	siempre
27	etoi153	Descargada/ingresada y en recepción (DM, papel) del encuestador.	siempre
30	etoi153	Disponible para ser asignada por el recepcionista.	nunca
32	etoi153	Asignada a recuperador.	nunca
33	etoi153	Cargada en el (DM); en campo para recuperador.	descargando
34	etoi153	Cargada en el (papel); en campo para recuperador.	nunca
35	etoi153	Descargada vacía para ingreso o para volver a cargar	nunca
36	etoi153	En ingreso (de recuperador)	siempre
37	etoi153	Descargada y en recepción (DM, papel) del recuperador.	siempre

asignados semanalmente a los encuestadores. Los lotes se cargan en un dispositivo móvil determinado, junto a su hoja de ruta digital, o se marcan para ser relevados en papel (sin dispositivo móvil). En ambos casos, se entrega la hoja de ruta impresa. Las viviendas que el encuestador no visitó pueden volver a entregarse a la semana siguiente.

Del mismo modo se realiza la asignación a los demás roles de campo: al recuperador (que visita las viviendas encuestables, rechazos o ausencias) o al supervisor (Figura 3).⁴

La recepción de encuestas

Cuando las encuestas son en papel, el rol principal del recepcionista es hacer un primer control de las encuestas que recibe del encuestador. Para ello, abre cada uno de los cuestionarios y sigue el flujo del mismo prestando atención a los pases, que no

⁴ Véanse los roles y tareas en el Anexo.

se hayan saltado bloques de preguntas, que no se hayan llenado de más, que las respuestas de texto de las preguntas que se utilizan para codificar la tarea de la población económicamente activa y la rama de actividad de las empresas donde trabajan (o de su actividad como cuentapropista) sean adecuadas y completas. También revisa manualmente las pautas de consistencia principales. En caso de encontrar problemas, hará las correcciones que sean pertinentes, advertirá al encuestador o iniciará el proceso de capacitación del encuestador.

Con dispositivos móviles, el encuestador ya no comete errores de omisión de bloques por distracción, pero todavía hay que prestar atención a los textos y a las inconsistencias. En ciertos contextos, escribir texto en el dispositivo móvil no es tan cómodo como con el papel, y eso hace que el encuestador se vea tentado a abreviar o escribir menos de lo necesario. La ventaja de utilizar el SIEH es que los controles de flujo y de consistencias se hacen en forma automática pudiendo correrse el lote completo de pautas de consistencias,

con lo que aumenta la calidad de la información al haber pasado por la tarea de recepción.

El recepcionista revisa el listado de omisiones y de inconsistencia en la pantalla de su puesto de trabajo con el encuestador presente. En caso de detectar alguna inconsistencia, se la informa al encuestador para que este explique si repreguntó o indagó con el encuestado y deje registrado en el sistema la justificación de las inconsistencias.

La revisión de los textos para la codificación de rama de actividad y tarea debe realizarse manualmente. El recepcionista completa las abreviaturas y los textos en caso de que el encuestador recuerde algún dato no escrito. En la DGEyC la recepción de las encuestas a hogares se realiza semanalmente con una carga de encuestas de entre 10 y 20 encuestas por lote (según la experiencia del encuestador), *de modo de que sea posible que el encuestador aún recuerde la entrevista y se puedan hacer correcciones sobre la base de la información provista por el hogar.*

Figura 4
Sistema Integrado de Encuestas a Hogares (SIEH): Control de cobertura y efectividad de campo

	total gral	total	todavía no salieron	no salen a campo	total salieron	en campo sin resultado	con resultado	no encuestables	encuestables/ausencias	contactados	rechazo	otras causas	rechazo otras causas	respondientes	otras condiciones	tasa nenc	tasa ausencia	tasa rechazo y otros	tasa resp efectiva	tasa efectividad	
Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	
F																					
	TOTAL	3190	510	0	2680	250	2430	356	2074	361	1713	244	15	259	1454	0	14.7	17.4	12.5	59.8	70.1
x	Grilla: vista_monitoreo_resumen_porea_semanal registros: 12																				
	semana	total	todavía no salieron	no salen a campo	total salieron	en campo sin resultado	con resultado	no encuestables	encuestables/ausencias	contactados	rechazo	otras causas	rechazo otras causas	respondientes	otras condiciones	tasa nenc	tasa ausencia	tasa rechazo y otros	tasa resp efectiva	tasa efectividad	
Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	
F																					
	1	260	0	0	260	0	260	42	218	41	177	19	0	158	0	16.2	18.8	8.7	60.8	72.5	
	2	270	0	0	270	0	270	43	227	35	192	16	1	175	0	15.9	15.4	7.5	64.8	77.1	
	3	280	0	0	280	0	280	38	242	40	202	19	0	183	0	13.6	16.5	7.9	65.4	75.6	
	4	280	0	0	280	0	280	47	233	28	205	27	0	178	0	16.8	12.0	11.6	63.6	76.4	
	5	260	0	0	260	0	260	34	226	42	184	21	1	162	0	13.1	18.6	9.7	62.3	71.7	
	6	270	0	0	270	0	270	58	212	26	186	17	1	168	0	21.5	12.3	8.5	62.2	79.2	
	7	270	0	0	270	0	270	36	234	44	190	27	4	159	0	13.3	18.8	13.2	58.9	67.9	
	8	270	0	0	270	0	270	30	240	49	191	42	3	146	0	11.1	20.4	16.8	54.1	60.8	
	9	270	0	0	270	0	270	28	242	56	186	56	5	125	0	10.4	23.1	25.2	46.3	51.7	
	10	260	10	0	250	250	0	0	0	0	0	0	0	0	0						
	11	250	250	0	0	0	0	0	0	0	0	0	0	0	0						
	12	250	250	0	0	0	0	0	0	0	0	0	0	0	0						

El monitoreo de campo: la cobertura y efectividad de campo

Una herramienta necesaria para la coordinación operativa de un proyecto estadístico es el monitoreo de su cobertura en función del calendario del operativo y la población muestral relevada necesaria para poder inferir los aspectos que se desea estudiar.

El módulo de monitoreo del SIEH (Figura 4) brindó la posibilidad de obtener en tiempo real las tablas descriptivas para analizar el resultado de las encuestas ingresadas al sistema (encuestas que salieron a campo y regresaron siendo realizadas o no realizadas y encuestas que no salieron a campo). A partir de estos datos, las tablas también incluyen indicadores tales como el porcentaje de encuestas efectivas, el porcentaje de encuestas no realizadas por motivo (vivienda sin hogar, vivienda con hogar ausente, rechazo del hogar en la vivienda, etc.), tasa de efectividad, tasa de rechazo y error de marco. Las tablas presentan los indicadores por semana, zona, comuna y encuestador.

El control de calidad durante el operativo

Las estimaciones de encuestas por muestreo se encuentran afectadas por errores muestrales y no muestrales. Los errores muestrales dependen del diseño elegido, estimadores, pruebas de significatividad, etc., que son aspectos medibles y controlables. Los errores no muestrales, según Naciones Unidas (2009), pueden provenir: del *marco muestral*, cuando hay elementos a empadronar que están omitidos, duplicados o cuando hay elementos erróneamente incluidos;

de la *falta de respuesta*, cuando un elemento de la muestra no responde o cuando el cuestionario se cumplimenta parcialmente; *del diseño conceptual del cuestionario*, cuando las preguntas, debido a conceptos o definiciones, están mal especificadas o expresadas; *de medida*, cuando los respondientes dan información incorrecta (intencionalmente o no) y/o cuando los encuestadores falsifican datos, inducen respuestas o registran incorrectamente; y *de procesamiento*, errores de grabación, codificación, asignación de ponderaciones y tabulación de los datos, entre otros.

Como ya se vino explicando en apartados anteriores, el SIEH contribuye a disminuir y controlar los errores no muestrales. En el caso de la supervisión, se obtuvo una aproximación a la medición de los errores de medida a partir de su incorporación al sistema. La modalidad de la supervisión que se aplica es la de encuesta repetida. La encuesta repetida consiste en que el supervisor vuelva a tomar contacto con el entrevistado para que reitere el cuestionario en una versión reducida. El procedimiento aplicado reproduce exactamente todas las pautas del proceso de la encuesta original y se aplica a una sola persona en el hogar, el respondiente de la encuesta. Posteriormente, seorean las respuestas obtenidas por el supervisor y por el encuestador y se arma, para cada pregunta, la tabla comparativa que en su diagonal contiene los resultados concordantes. El objetivo es calcular la diferencia entre los valores de las variables obtenidos en las dos entrevistas (Figura 5).

La encuesta repetida (supervisión) se releva utilizando dispositivo móvil y su circuito es análogo al del resto de las encuestas.

Figura 5
Sistema Integrado de Encuestas a Hogares (SIEH): Módulo seguimiento de encuestas

	semana	enc	area	comuna	nombre	hn	hp	hd	hab	manzana	casa	barrio	ident edif	obs	estado	repcionista	cod enc	fecha carga enc	fec primet
ir	1	100707	1007	4	OLAVARRIA	119	9	A							41	220	73		
ir	1	103604	1036	9	PILAR	266									41	241	130	7/10/2014	7/10/14
ir	2	205408	2054	1	POSADAS	2420	2	A					1		41	251	178	14/10/2014	14/10/14
ir	2	205610	2056	1	LAVALLE	3845	11	POR							41	240	95	14/10/2014	14/10/14
ir	1	300701	3007	2	AV CORDOBA	8709	9	A							41	207	91	7/10/2014	7/10/14
ir	1	302703	3027	7	AV JUAN BAUTISTA ALBERDI	1103	0	A							41	211	85	6/10/2014	6/10/14
ir	1	303201	3032	8	AV LARRAZABAL	1771									41	256	165	6/10/2014	6/10/14
ir	1	303301	3033	8	GRAL NAPOLEON URIBURU	4395	5	00D				BARRIO SAVIO	TIRA C BLOCK 26		41	201	49	6/10/2014	6/10/14
ir	1	303902	3039	10	AV SEGUROLA	4155	7	D							41	241	116	6/10/2014	6/10/14
ir	1	303907	3039	10	DR DAVID	5675	0	B00							41	241	116	6/10/2014	6/10/14

El procesamiento

El procesamiento de las encuestas a hogares comprende las tareas de revisión y consistencia de los datos ingresados, la codificación de variables abiertas, el control de datos agrupados, la creación de variables, la imputación de variables omitidas (en particular, la imputación de ingresos monetarios), los cálculos de factor de expansión y calibración, la generación de tabulados y la exportación de base a usuario y bases internas de trabajo.

El SIEH contiene un conjunto de herramientas para desempeñar las tareas de procesamiento.

Control de inconsistencias

Durante el procesamiento deben evaluarse las inconsistencias detectadas para decidir la validez del dato (para las pautas de consistencia de tipo advertencia) o la acción correctiva según el caso.

Como ya se dijo, las inconsistencias son mostradas al encuestador durante el relevamiento para que pueda corregir errores de marcado o repreguntar para obtener más información. Las inconsistencias no corregidas se le muestran al recepcionista que analiza junto al encuestador, corrige las que correspondan y registra las aclaraciones para cada una de las inconsistencias que no puede –o considera que no debe– corregir.

En la primera etapa de procesamiento, se evalúan todas las inconsistencias detectadas analizando los datos en su conjunto y utilizando las observaciones del encuestador y las aclaraciones del recepcionista.

Las pautas de consistencia son controladas periódicamente sobre la base de obtener frecuencias de ocurrencia y comparándose con operativos anteriores. De estas frecuencias, se puede detectar la necesidad de capacitar a algún encuestador sobre algún tema puntual.

Codificación y reclasificación de variables abiertas

Las variables abiertas deben codificarse para poder utilizarse en la construcción de otras variables, por ejemplo, el país de nacimiento. Otros textos que deben analizarse son los “otros especificar” que corresponden a las variables cerradas donde se controla si algunos textos ingresados corresponden a categorías existentes y deben ser reclasificados o quedan en la categoría otros.

El SIEH mantiene un conjunto de diccionarios de textos que traducen las distintas formas de escribir ciertos valores al clasificador utilizado (por ejemplo, nombres de países a código de país o nombres de calles o barrios o sus respectivos códigos). Estos diccionarios se inician con los textos estándar y se actualizan con las distintas formas de expresiones textuales que fueron apareciendo en cada operativo.

Los diccionarios se utilizan para el proceso asistido de codificación y clasificación de variables. Cuando un texto no puede codificarse en forma automática, un especialista en el tema la clasifica manualmente. Esta clasificación es incluida en el diccionario referido para el uso en el operativo actual y en los subsiguientes.

La codificación de la rama de actividad del lugar de trabajo y de la ocupación de la persona es un proceso en el que intervienen varias variables en forma simultánea; la clasificación se hace dígito a dígito y tiene un componente manual importante. El sistema despliega en una grilla (Figura 6) un renglón por cada caso a codificar y en las columnas las variables necesarias para la tarea; cada dígito a codificar está en una columna. El técnico que codifica puede filtrar u ordenar la grilla por cualquiera de las variables para codificar en lotes de información similar. Al finalizar la codificación, la información es revisada por un supervisor de codificación que utiliza las mismas grillas de codificación.

Construcción de variables

Los resultados de la encuesta incluyen tabulados que están formados por variables que no son la respuesta directa a una pregunta del cuestionario, sino que se arman utilizando una combinación de ellas (por ejemplo, la condición de actividad o los años de escolaridad de una persona). La construcción de variables la hace automáticamente el sistema sobre la base de la carga de las fórmulas que las definen. Tener las variables automatizadas dentro del sistema integrado permite usarlas en las pautas de consistencias y en la generación de bases usuario.

Figura 6
Sistema Integrado de Encuestas a Hogares (SIEH): Módulo de codificación de rama y tarea

Grilla: 11_rama_ocupacion, registros: 4949																	
cond	catgori	rama1	rama2	137	137sd	138	139	ocu1	ocu2	ocu3	ocu4	ocu5	141 cod2	140a	141	142	143
1	2	32	00	3200		2	2	8	0	1	1	3	80113	1	Orfebre, artesano	Hace joyas y las vende en puesto en feria	Martillo, limas, pinzas.
1	3	82	00	8200		2	2	3	0	3	3	3	30333	4	Operadora	Vende productos y servicios por teléfono	Teléfono, computadora
1	3	48	03	4803		2	2	3	0	3	1	4	30314	1	Empleado quiosco	Atiende al público en quiosco	Ninguna
1	3	97	00	9700	1			5	7	3	1	4	57314				
1	3	61	00	6100		2	2	1	0	3	3	3	10333	4	Recepcionista	Tareas administrativas, gestión de envíos de tramites, manejo de conmutador	Teléfono de línea con conmutador, computadora.
1	3	35	01	3501		2	3	1	0	2	0	3	10203	4	Coordinadora de servicios	Coordinar horarios y personal para el servicio de	Computadora, teléfonos y vehículo

Las variables también son usadas en combinación con los diccionarios y sirven al proceso asistido de codificación.

El módulo del sistema que permite la definición de las variables (Figura 7) provee cuadros de frecuencias de cada uno de los casos que definen los valores de la nueva variable para poder controlar la construcción de la misma (los casos se definen por las distintas combinaciones de las variables utilizadas como insumos –por ejemplo, para ocupados son distintos casos una persona que trabajó en la semana de referencia y una persona que no trabajó pero tenía un trabajo al que no concurrió porque estaba de vacaciones–). Esos cuadros de frecuencia se controlan con operativos anteriores.

Para la creación de variables, se pueden escribir fórmulas que combinen variables de la misma unidad de análisis (por ejemplo: edad y condición de actividad de una persona) y de una unidad de análisis que la incluya (por ejemplo: para construir variables de personas se pueden utilizar variables de hogar o de vivienda), o variables que agrupen otras de unidades de análisis incluidas (por ejemplo: en la tabla de hogares se puede agregar la variable cantidad de miembros menores de 18 años).

En la creación de variables, es importante considerar el orden en que se crean porque ciertas variables construidas pueden ser insumo de otras variables construidas.

Figura 7 Sistema Integrado de Encuestas a Hogares (SIEH): Módulo de definición de variables calculadas

Grilla: varcal, registros: 4											
varcal	destino	orden	nombre	activa	baseusuario	nombrevar	baseusuario	tipodedato	grupo	valida	
B	C	cond_activ	mie 580	Condición de actividad	Sí	Sí	estado	entero	trabajo	Sí	
Grilla: varcalopc											
opcion	expresion condicion							etiqueta			
B	1	t_ocup >= 1							Ocupado		
B	2	t_desoc >= 1							Desocupado		
B	3	t_ina >= 1							Inactivo		
B	C	t_desoc	mie 370	t desoc	Sí	no		entero	trabajo	Sí	
B	C	t_ina	mie 530	t ina	Sí	no		entero	trabajo	Sí	
B	C	t_ocup	mie 290	T ocup	Sí	no		entero	trabajo	Sí	
Grilla: varcalopc											
opcion	expresion condicion							etiqueta			
B	-9	t1<0 or t2<0 or t3<0 or t4<0 or t5<0 or t6<0 or t8<0							Ns/Nc		
B	1	t1=1 and t7=1							Trabajo con remuneración		
B	2	t1=1and t7=2 and (t8=1 or t8=2)							Trabajador familiar o ad-honorem		
B	3	t1=2 and t2=1 and t7=1							Changa con remuneración		
B	4	t1=2 and t2=1 and t7=2 and (t8=1 or t8=2)							Changa para familiar o ad-honorem		
B	5	t1=2 and t2=2 and t3=5 and (t4>=1 and t4<=3)							Con licencia laboral o vacaciones		
B	6	t1=2 and t2=2 and t3=5 and t4=4 and t5=1							Suspensión con pago		
B	7	t1=2 and t2=2 and t3=5 and t4=5 and t6=1							Con trabajo asegurado		

Cálculo del factor de expansión y calibración

Las encuestas a hogares son expandidas y calibradas por el Departamento de Muestreo utilizando las técnicas estadísticas correspondientes en función del diseño muestral. En el caso de la EAH, el diseño y los tamaños de muestra están preparados para dar información desagregada por comuna. En el caso de la ETOI, los resultados se desagregan por zona (norte, centro y sur).

Los factores de expansión y calibración son incorporados al SIEH y utilizados para la producción de los tabulados. Los tabulados se pueden generar muestrales (sin considerar los factores de expansión); estos se usan durante el procesamiento en las etapas de control. Para calcular los factores de expansión y calibración, tiene que estar terminado de cerrar el trabajo de campo y estar determinados los códigos de no respuesta (en particular, diferenciadas las viviendas no encuestables de las encuestables).

El factor de expansión y calibración se exporta junto con la base usuario y bases de trabajo.

Generación de tabulados

El resultado de las encuestas se resume en cuadros o tabulados (Figura 8) que, en el caso de las encuestas periódicas como la EAH y la ETOI, se publican periódicamente sobre la base de un calendario y un plan de publicación. Es imprescindible, para no perder la comparabilidad, que esos cuadros se generen utilizando siempre las mismas definiciones.⁵

Las definiciones de los cuadros son parte de los metadatos del sistema: nombre del cuadro, variables en filas y columnas, filtro aplicado y tipo de cuadro (tasas, valores absolutos, etc.).

⁵ A veces alguna definición puede cambiar en algún detalle por el agregado o supresión de alguna pregunta o de alguna categoría dentro de las opciones de alguna pregunta. En estos casos, se publica junto al cuadro las aclaraciones pertinentes.

Figura 8

Sistema Integrado de Encuestas a Hogares (SIEH): Tabulado de resultados

Cuadro 15 Distribución porcentual de la población de 3 años y más que asiste a un establecimiento educativo por sector de gestión al que pertenece el establecimiento según comuna. Ciudad de Buenos Aires. Año 2016

Comuna	Sector de gestión			
	Total	Estatal/ público	Privado	Ns/Nc
Total	100,0	56,8	43,2	0,0
1	100,0	72,5	27,5	-
2	100,0	43,8	56,2	-
3	100,0	54,8	45,2	-
4	100,0	77,1	22,9	-
5	100,0	54,9	45,1	-
6	100,0	52,8	47,2	-
7	100,0	64,4	35,6	-
8	100,0	70,6	29,4	-
9	100,0	59,9	40,1	-
10	100,0	56,6	43,4	-
11	100,0	57,8	42,2	-
12	100,0	50,8	49,2	-
13	100,0	36,1	63,9	-
14	100,0	32,7	67,1	0,2
15	100,0	56,0	44,0	-

Nota: el análisis fue realizado con la edad declarada. *Por tratarse de valores marginales, no se presentan los coeficientes de variación Ns/Nc.

Fuente: Dirección General de Estadística y Censos (Ministerio de Hacienda GCBA). EAH 2016.

Tener definidos los tabulados como metadatos permite definir nuevos cuadros para las tareas de control, nuevas publicaciones o procesamientos para ocasiones especiales (Figura 9).

El sistema permite generar los tabulados con datos muestrales o expandidos, con el coeficiente de variación en números o letras.⁶

Exportación de base a usuarios y bases internas de trabajo

El SIEH tiene un módulo de exportación de bases (Figura 10) donde se define la selección de variables, su orden, nombre y destino (hogares o miembros del hogar). De ese modo, se puede poner variables del hogar en la tabla de miembros (por simplicidad una de las variables que se pone en ambos destinos es el factor de expansión y calibración).

⁶ Se generan las celdas de los tabulados marcadas con "a" cuando el coeficiente de variación está entre 10 y 20% y marcadas con "b" para los mayores al 20%

Figura 9
Sistema Integrado de Encuestas a Hogares (SIEH): Módulo de definición de tabulados

Grilla: tabulados , registros: 9								
a)	tab	titulo	fila1	fila2	columna	cel exp	cel tipo	filtro
B	CUADRO 1	Distribución porcentual de la población por sexo según comuna.	comuna		sexo	personas	comun	
B	CUADRO 2	Distribución porcentual de la población por grupo de edad según comuna.	comuna		edad10a	personas	comun	
B	CUADRO 3	Distribución porcentual de la población por grupo de edad según sexo y zona.	sexo	zona_3	edad10a	personas	comun	
B	CUADRO 4	Distribución porcentual de los jefes de hogar por grupo de edad según comuna.	comuna		edad10b	personas	comun	p4=1
B	CUADRO 5	Distribución porcentual de los jefes de hogar por condición de actividad según comuna.	comuna		cond_activ	personas	comun	p4=1
B	CUADRO 6	Distribución porcentual de los jefes de hogar por sexo según comuna.	comuna		sexo	personas	comun	p4=1
B	CUADRO 7	Distribución porcentual de los jefes de hogar por grupo de edad según sexo y zona.	sexo	zona_3	edad10b	personas	comun	p4=1
B	CUADRO 8	Distribución porcentual de los hogares por cantidad de personas en el hogar y tamaño medio según zona.	zona_3			hogares	comun	
B	CUADRO 9	Distribución porcentual de los hogares por régimen de tenencia de la vivienda según comuna.	comuna		h2_re	hogares	comun	

Figura 10
Sistema Integrado de Encuestas a Hogares (SIEH): Módulo de definición de bases a exportar

Grilla: baspro , registros: 31										
a)	ope	baspro	nombre			cambiar especiales	cambiar nsnc por	cambiar sindato por	cambiar null por	
B	eah2017	base_dh	base dh			Sí	9	0	0	
B	eah2017	base_estruc	base plan uso interno			no				
B	eah2017	bu_ampliada	Base Usuarios Ampliada EAH			Sí	9	0	0	
			var	orden	alias	exportar en	for	destino	largo para documentacion	
			enc_bu	110	id	ambas		tem	Encuesta Base Usuario	
			hogar_bu	120	nhogar	ambas		hog	Hogar	
			mie_bu	130	miembro			mie	Número de miembro para Base usuarios	
			comuna	170		ambas				
			dominio_bu	180	dominio	ambas		hog	Dominio	
			v2	200	v2_2		A1		Tipo de vivienda	
			v4	210			A1		¿Cuántas habitaciones/ ambientes tiene, en total, esta vivienda? Total de habitaciones/ambientes	
			h2_re_bu	240	h2			hog	Régimen de tenencia de la vivienda	
			h3	245			A1		¿Cuántas habitaciones/ ambientes son de uso exclusivo de este hogar?	
			hacinam_2	360				hog	Hacinamiento	

En el módulo de exportación, se define la base usuarios (y de ese modo se generan siempre las mismas columnas en el mismo orden) y las bases internas de trabajo, que son las utilizadas para extraer información que es insumo de otras áreas de la DGEyC no incorporadas al SIEH (la garantía de formato y orden facilita la automatización e integración con otros sistemas informáticos).

La exportación genera un único archivo comprimido que contiene un archivo de datos por cada destino (hogar y miembros) y sus diseños de registros.

Proyectos estadísticos que promovieron ampliar el horizonte

El SIEH se diseñó para ser utilizado en la EAH. Si bien el dispositivo móvil se empezó a usar en la EAH 2012, el SIEH comenzó a emplearse para el ingreso y la primera etapa de procesamiento (que consiste en el control de las encuestas ingresadas basadas en el control de pautas de consistencia) de la EAH 2011. En las sucesivas EAH, se fueron incorporando módulos: procesamiento, supervisión, generación de bases usuarios, etcétera.

El SIEH se utilizó para todas las encuestas producidas por la Subdirección General Sociodemográfica:

- 2012, Encuesta de Acceso a la Justicia
- 2013 y continúa anualmente, Encuesta sobre el perfil de los Asistentes de la Feria Internacional del Libro de Buenos Aires
- 2014, Encuesta Epidemiológica de Salud Mental de la Ciudad de Buenos Aires (EESAME 2014)
- 2014 hasta la actualidad, Encuesta Trimestral de Ocupación e Ingresos (ETOI)

- 2014, Encuesta a Beneficiarios de Ciudadanía Porteña
- 2015, Encuesta de Abonados del Teatro Colón
- 2015, Prueba Piloto Encuesta de Uso del Tiempo de la Ciudad de Buenos Aires
- 2016, Encuesta de Uso del Tiempo de la Ciudad de Buenos Aires (UT CABA 2016)
- 2016, Empadronamiento en Villa 31 y 31 bis (bajo autopista)
- 2017, Prueba Piloto Encuesta Demográfica Retrospectiva de la Ciudad de Buenos Aires
- 2017, Empadronamiento en Barrio 31 y 31 bis
- 2017, Relevamiento de Personas en Situación de Calle

UT CABA 2016 El diario de actividades

La encuesta de Uso del Tiempo requirió de un desarrollo especial para el registro del diario de actividad, dado que no se ajustaba al modelo clásico de pregunta/respuesta.

El objetivo de la encuesta de Uso del Tiempo es preguntarle al encuestado todo lo que hizo durante el día anterior al relevamiento y cuánto tiempo le dedicó a la actividad: dormir, trabajar, viajar, cuidar niños, cocinar, estudiar, hacer deportes, realizar tareas domésticas, etc. Algunas de esas actividades pueden hacerse en forma simultánea, algunas pueden parecer más importantes o son más fáciles de recordar que otras. Para los objetivos de la encuesta, todas las actividades son igualmente importantes, y en todo momento del día tiene que haber al menos una actividad registrada (aunque sea descansar).

De acuerdo con la prueba piloto⁷ realizada en 2015, la forma más adecuada de llegar al objetivo es preguntarle al encuestador qué fue haciendo durante el día. Simplificando el ejemplo, la encuesta fue más o menos así:

Hablemos del día de ayer, ¿a las 0hs del día de ayer usted qué estaba haciendo? [a/r] ¿Hasta qué hora realizó esa actividad? [a/r] ¿Realizó en ese horario alguna otra actividad, desde qué hora a qué hora? [a/r] ¿Alguna otra? [a/r].

Si, por ejemplo la respuesta fue que a las 0hs estaba durmiendo y durmió hasta las 6:30, la encuesta continúa preguntando desde esa hora:

Continuemos, ¿qué hizo a las 6:30? [a/r] ¿Hasta qué hora realizó esa actividad? [a/r] ¿Realizó en ese horario alguna otra actividad, desde qué hora a qué hora? [a/r] ¿Alguna otra? [a/r].

Y continúa de ese modo. El módulo "Diario de actividad" de la de la EAH 2005 era una matriz en papel con 3 columnas (suponiendo un máximo de 3 actividades simultáneas) y un renglón cada media hora. Este formato de diario permite fácilmente registrar las actividades que vaya recordando el encuestado aun cuando recuerde algo que hizo a la mañana mientras contesta las cosas de la tarde.

7 En 2005, se había realizado la encuesta sobre la base de un diario de actividades en una encuesta en papel. En 2015, se hizo una prueba piloto de una encuesta en la que, en vez de preguntar en forma secuencial qué hizo durante el día, se preguntó cuántas horas le había dedicado el día anterior a una lista de actividades. La conclusión de la prueba piloto fue que lo más adecuado era preguntar secuencialmente con un "diario de actividades". Para la encuesta definitiva de 2016, se decidió utilizar un diario de actividades ayudado con una tarjeta donde estuvieran listadas todas las actividades; además, esa tarjeta estaba ilustrada y codificada de modo de facilitar la recordación del encuestado y el llenado del diario por el encuestador que debía consignar el código de actividad.

La encuesta en 2016⁸ se llevó a cabo en dispositivo móvil; se realizó un desarrollo especial para el llenado del diario de actividades (Figura 11). En una serie de tiempo gráfica de 3 columnas, mostraba para cada actividad un rectángulo de tamaño proporcional al tiempo usado que se dibujaba sobre las marcas de horas correspondientes. Las 3 columnas permitían mostrar gráficamente las actividades realizadas con simultaneidad. El programa validaba códigos y horarios, permitía registrar actividades recordadas en momentos anteriores, graficaba la línea de tiempo a medida que se iba llenando el cuestionario y destacaba gráficamente sobre la línea de tiempo los períodos en los que no se había informado ninguna actividad.

El resto de las preguntas de los cuestionarios eran similares a las de las demás encuestas a hogares.

Para el procesamiento se agrega la *unidad de análisis episodio* que corresponde a un período de tiempo durante el cual se realiza determinada actividad; se calcula el tiempo total (o sea el tiempo transcurrido en el reloj para ese episodio) y el tiempo con simultaneidad (que corresponde al prorrateo del tiempo entre otros episodios simultáneos). La suma del tiempo con simultaneidad de todos los episodios es 24hs, la suma del tiempo total podría ser mayor. Esas dos variables calculadas se utilizan como insumo en la mayoría de los tabulados de resultados y se incluyen en la base usuario.

8 En 2016 se realizó una encuesta específica de Uso del Tiempo, mientras que en 2005 se aplicó un módulo a la EAH.

Figura 11
Sistema Integrado de Encuestas a Hogares (SIEH): Diario de Actividades. Registro de hasta 3 actividades simultáneas

H	M	Act1	Act2	Act3	H	M	Act1	Act2	Act3	desde	hasta	c.act.	R	observaciones
0	00	922			12	10	1			00:00	07:30	922		Dormir
1	25	Dormir			13	20	Trabaja			07:30	08:00	921		Comer y beber
2	40				14	30	para un			08:00	09:00	911		Higiene, arreglo personal, etc.
3	00				15	00	patrón			09:00	10:00	14		Viajar para ir y volver al trabajo
4	15				16	10	o por			10:00	18:00	1		Trabajar para un patrón o por cu
5	30				17	20	cuenta			18:00	19:10	14		Viajar para ir y volver al trabajo
6	45				18	00	propia			19:10	20:00	36		Hacer compras para el hogar
7	00				19	10				20:00	21:30	61		Estudiar
8	15	921			20	20	14			20:00	21:00	411		Cuidado personal a niños del hc
9	30	911			21	30	36			21:30	22:30	31		Preparar y servir la comida
10	45	14			22	00	61	411		22:30	23:30	82		Ver televisión
11	00	1			23	10	Estudia			23:30	24:00	911		Higiene, arreglo personal, etc.
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
	30					30								
	45					40								
	00					50								
	15					00								
	30					10								
	45					20								
	00					30								
	15					40								
	30					50								
	45					00								
	00					10								
	15					20								
</														

ción de inconsistencias dentro de las matrices y el control de completitud.

Es necesario *definir nuevas formas de procesamiento* y, para ello, se deben desarrollar módulos en el sistema informático: además de agregarse la dimensión tiempo a las unidades de análisis es necesario "completar las bases" generando un registro por año para cada una de ellas (años-vividos de la persona encuestada).

Estadísticas de población sobre la base de datos de registro

La experiencia acumulada en la captación y procesamiento de las encuestas a hogares, en especial el tratamiento de las mismas basado en metadatos, permite extrapolar los sistemas desarrollados a tareas estadísticas que no estén basados en la realización de encuestas.

La DGEyC tiene una vasta experiencia en la consolidación, sistematización y producción estadística basada en información provista por proveedores externos; en este aspecto se destacan los registros administrativos. Esta información está estructurada en unidades de análisis, cada una con sus variables bien definidas. La información se obtiene a partir de cronogramas acordados y es revisada antes de consolidarse en las oficinas de la DGEyC. Estas estadísticas están basadas en datos del Registro Civil, de escuelas de la Ciudad, de hospitales, etc. Cada fuente se procesa de acuerdo con un procedimiento específico.

Se plantea como desafío extender el sistema para incluir las tareas de la producción estadística de población basada en datos de registros, de modo de unificar los procesos utilizando los mismos módulos de consolidación cada vez que sea posible, para compartir diccionarios, definiciones de variables y metodología de trabajo. De este modo, se puede avanzar en la interconexión digital con las áreas productoras de información armonizando las definiciones conceptuales.

Conclusiones

El uso de metadatos en todos los aspectos del sistema tiene como consecuencia inmediata la documentación de la producción estadística. Esto aporta sistematicidad y transparencia a la producción de estadística pública.

El sistema informático se diseñó en simultáneo con el Sistema Integrado de Encuestas a Hogares. Participaron activamente en dicho diseño todos los equipos de trabajo: campo, logística, procesamiento, diseño conceptual, cartografía, muestreo, informáticos, especialistas temáticos y sectorialistas. La metodología de trabajo estuvo guiada por el Diseño de Experiencia de Usuario, la Teoría de Usabilidad y lo que actualmente se llama Diseño de Servicios.

La implementación del Sistema Integrado de Encuestas a Hogares permitió aumentar y mejorar la producción de estadísticas sociodemográficas de la Dirección General de Estadística y

Censos de la Ciudad de Buenos Aires.

A.1

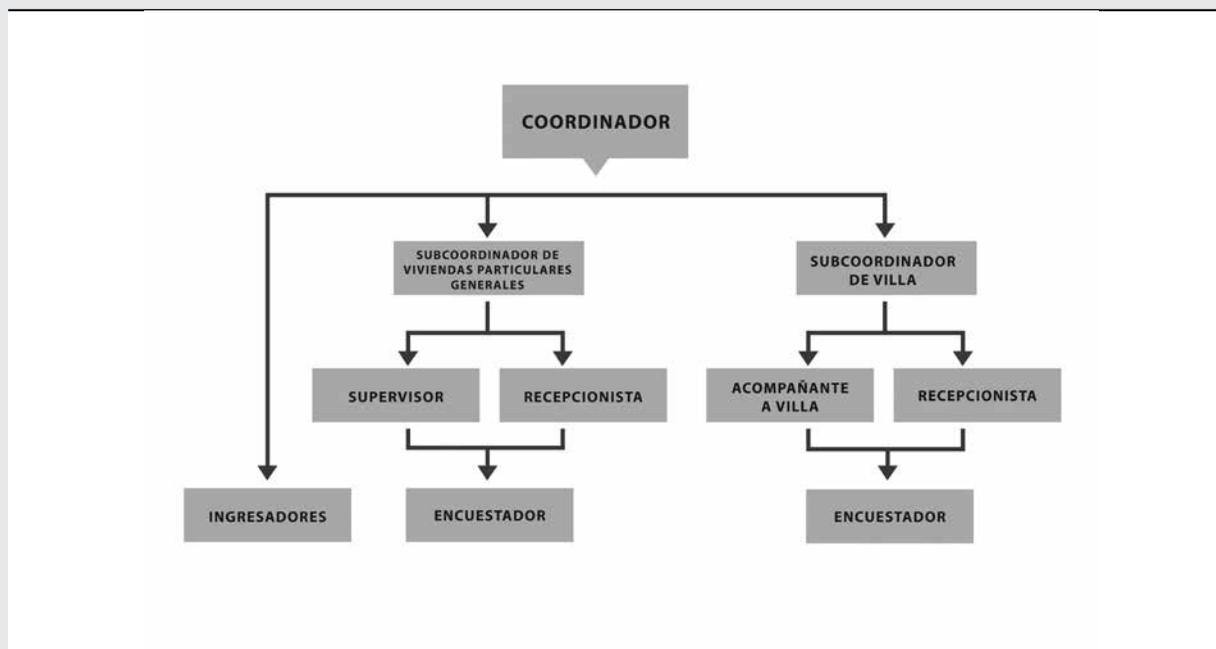
Calendarios de semanas

Grilla: **semanas**, registros: 12

sem	semana referencia desde	semana referencia hasta	30días referencia desde	30días referencia hasta	mes referencia	carga enc desde	carga enc hasta	carga recu desde	carga recu hasta
$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$	$\Delta\ominus$
1	2017-12-26	2018-01-01	2017-12-03	2018-01-01	2017-12-01	2017-12-29	2018-01-19	2018-01-11	2018-01-28
2	2017-12-31	2018-01-06	2017-12-08	2018-01-06	2017-12-01	2018-01-08	2018-01-24	2018-01-17	2018-02-02
3	2018-01-07	2018-01-13	2017-12-15	2018-01-13	2017-12-01	2018-01-15	2018-01-31	2018-01-24	2018-02-09
4	2018-01-14	2018-01-20	2017-12-22	2018-01-20	2017-12-01	2018-01-22	2018-02-07	2018-01-31	2018-02-16
5	2018-01-21	2018-01-27	2017-12-29	2018-01-27	2018-01-01	2018-01-29	2018-02-14	2018-02-07	2018-02-23
6	2018-01-28	2018-02-03	2018-01-05	2018-02-03	2018-01-01	2018-02-05	2018-02-21	2018-02-14	2018-03-02
7	2018-02-04	2018-02-10	2018-01-12	2018-02-10	2018-01-01	2018-02-12	2018-02-28	2018-02-21	2018-03-09
8	2018-02-11	2018-02-17	2018-01-19	2018-02-17	2018-01-01	2018-02-19	2018-03-07	2018-02-28	2018-03-16
9	2018-02-18	2018-02-24	2018-01-26	2018-02-24	2018-01-01	2018-02-26	2018-03-14	2018-03-07	2018-03-23
10	2018-02-25	2018-03-03	2018-02-02	2018-03-03	2018-02-01	2018-03-05	2018-03-21	2018-03-14	2018-03-30
11	2018-03-04	2018-03-10	2018-02-09	2018-03-10	2018-02-01	2018-03-12	2018-03-28	2018-03-21	2018-04-06
12	2018-03-11	2018-03-17	2018-02-16	2018-03-17	2018-02-01	2018-03-19	2018-04-04	2018-03-28	2018-04-13

A.2

Estructura operativa de campo. Esquema 1. Estructura operativa de la Encuesta Anual de Hogares



A.3

Estructura operativa de campo. Puesto, función y recursos humanos de la estructura operativa básica de las encuestas a hogares

Puesto	Función
Coordinador	Arbitrar los medios y definir las estrategias para la consecución de la encuesta (elaboración del cronograma de trabajo, actividades y fechas, estructura organizativa y presupuesto del operativo de campo, requerimientos de recursos humanos; planificación de la capacitación; documentación del personal a contratar y remisión al sector responsable de las contrataciones; previsión de la impresión de los materiales; carga de trabajo y asignación de las mismas y demás especificidades del operativo).
Subcoordinador	Asistir a la coordinación, distribuir los recursos humanos por comuna, adjudicar la carga de trabajo al recepcionista y al encuestador, cargar en el sistema informático y controlar la supervisión presencial y telefónica. Monitorizar el operativo en las comunas a su cargo y el trabajo del recepcionista.
Recepcionista	Adjudicar la carga de trabajo al encuestador, recibir las encuestas en papel para entregar al <i>data entry</i> y recibir las encuestas en dispositivo móvil, editarlas, y realizar la supervisión telefónica.
Supervisor	Controlar los datos registrados en los cuestionarios relevados por el encuestador/recuperador en campo.
Recuperador	Revisitar las viviendas para lograr las entrevistas que, por diversos motivos, no pudo realizar el encuestador y completar el cuestionario.
Encuestador	Recibir la carga de trabajo asignada, ubicar las viviendas seleccionadas mediante la dirección que figura en su hoja de ruta, visitar las viviendas, realizar la entrevista y completar el cuestionario, entregar el trabajo al recepcionista.
<i>Data entry</i>	Ingresar al sistema informático las encuestas en papel en el momento de la recepción para su posterior edición, a cargo del recepcionista.

Fuente: Dirección General de Estadística y Censos (Ministerio de Economía y Finanzas GCBA).

Anexo

Bibliografía

Nielsen, Jakob (1989), *Coordinating User Interfaces for Consistency*, San Francisco, Morgan Kaufmann Publishers. [Edición de 2002].

————— (2012), *Mobile Usability*, Berkeley, New Riders Press.

Feres, Juan Carlos y Fernando Medina (2001), “Hacia un sistema integrado de encuestas de hogares en los países de América Latina”, en https://repositorio.cepal.org/bitstream/handle/11362/4707/1/S01010053_es.pdf.

Filgueira, C (2005), “Reflexiones acerca de los desafíos de la construcción de un sistema integrado de información estadística”, en <http://www.unfpa.org.uy/userfiles/informacion/items/293.pdf>.

Naciones Unidas (2009). *Diseño de muestras para encuestas a hogares: directrices prácticas. Estudio de métodos. Serie FNº 98*, Nueva York, Departamento de Asuntos Económicas y Sociales. División de Estadística.