

NOTAS DE POBLACIÓN

AÑO XXIX, N° 75, SANTIAGO DE CHILE



NACIONES UNIDAS



Comisión Económica para América Latina y el Caribe
Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de Población

LC/G. 2186-P
Diciembre de 2002

Copyright © Naciones Unidas 2003
Todos los derechos están reservados
Impreso en Chile

La autorización para reproducir total o parcialmente esta obra debe solicitarse al Secretario de la Junta de Publicaciones. Sede de las Naciones Unidas, N.Y.10017, EE.UU. Los Estados miembros y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Sólo se les solicita que mencionen la fuente e informen a las Naciones Unidas de tal reproducción.

PUBLICACIÓN DE LAS NACIONES UNIDAS

NÚMERO DE VENTA: S.03.II.G.77

ISBN 92-1-322063-4
ISSN impreso 0303-1829
ISSN electrónico 1681-0333

Ilustración de portada: Curdo Bermúdez, "El balcón" (detalle), 1941.
Gentileza del Banco Interamericano de Desarrollo
Diseño de portada: María Eugenia Urzúa

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE
José Antonio Ocampo Secretario Ejecutivo

CENTRO LATINOAMERICANO Y CARIBEÑO DE DEMOGRAFIA
(CELADE) – DIVISIÓN DE POBLACIÓN
Miguel Villa, Oficial a cargo

La Revista **NOTAS DE POBLACIÓN** es una publicación del Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población, cuyo propósito principal es la difusión de investigaciones y estudios de población sobre América Latina y el Caribe, aun cuando recibe con particular interés artículos de especialistas de fuera de la región y, en algunos casos, contribuciones que se refieren a otras regiones del mundo. Se publica dos veces al año (junio y diciembre), con una orientación interdisciplinaria, por lo que acoge tanto artículos sobre demografía propiamente tal, como otros que aborden las relaciones entre las tendencias demográficas y los fenómenos económicos, sociales y biológicos.

Comité editorial:

Jorge Bravo
Juan Chackiel
José Miguel Guzmán
Rolando Sánchez
Susana Schkolnik

Coordinador Técnico:

Juan Enrique Pemjean

Secretaria:

María Teresa Donoso

Redacción y administración:

Casilla 179-D, Santiago, Chile
E-mail: mdonoso@eclac.cl

Precio del ejemplar: US\$ 12

Suscripción anual: US\$ 20

Las opiniones expresadas en esta revista son responsabilidad de los autores, sin que el CELADE sea necesariamente participe de ellas.

SUMARIO

Presentación	7
Foreword	10
Avant-propos	13
América Latina: los censos del 2000 y el desarrollo social. <i>José Miguel Guzmán y Susana Schkolnik</i>	17
Los censos en América Latina: nuevos enfoques. <i>Juan Chackiel</i>	45
Cómo adaptar las nuevas tecnologías a las operaciones censales. <i>Arij Dekker</i>	73
Algunas reflexiones acerca de un formulario censal diseñado para la captura de datos mediante tecnología de imágenes (escáner). <i>Nelly T. Niedworok</i>	115
Los censos de población en Internet. <i>Dr. Werner Haug</i>	133
El censo de Canadá de 2006. Dirección estratégica. <i>Anil Arora</i> ..	147
¿Podrán los censos de los países en desarrollo responder a las futuras demandas de información?: el caso de Argentina. <i>Alejandro Giusti</i>	165
Preguntas para investigar la fecundidad y la mortalidad en la niñez y la mortalidad de adultos. <i>Ralph Hakkert</i>	195
La población indígena en Costa Rica según el censo del año 2000. <i>Elizabeth Solano Salazar</i>	217

CÓMO ADAPTAR LAS NUEVAS TECNOLOGÍAS A LAS OPERACIONES CENSALES *

Arij Dekker

División de Estadística, Secretaría de las Naciones Unidas

RESUMEN

Una mejora incluso pequeña en las tecnologías censales puede significar una ventaja significativa en la calidad o el costo de toda la operación. Actualmente, varias organizaciones están tratando de ayudar a incorporar innovaciones a las operaciones censales y estadísticas. Estas nuevas tecnologías plantean algunos problemas, como los siguientes: la elección de la tecnología apropiada; el mantenimiento de la integridad de los sistemas estadísticos existentes; la forma de abordar la externalización de determinadas tareas; y el mantenimiento de la confidencialidad de los datos. Algunas, como la telefonía móvil, han facilitado la comunicación interpersonal, del mismo modo que el fax y el correo electrónico. A su vez, la tecnología de los códigos de barras ha contribuido a mejorar la eficiencia de las actividades de gestión de los materiales utilizados en los censos.

En la ronda de censos de 2000, el reconocimiento inteligente de caracteres (RIC) constituyó un hito en varios países, aunque se observaron dificultades en la lectura de los caracteres manuscritos ilegibles y de los formularios impresos en forma deficiente. En general, las operaciones de los países que planificaron cuidadosamente el uso de las nuevas tecnologías

* Documento preparado para la Reunión del Grupo de Expertos del Simposio sobre el Examen Mundial de la Ronda de 2000 de los Censos de Población y Vivienda: Evaluación de mitad del decenio y perspectivas futuras (Nueva York, 7 al 10 de agosto de 2001).

y realizaron pruebas previas dieron mejores resultados. También se está ensayando el paso siguiente, a saber, la codificación automática o asistida por computadora; algunos datos, como los nombres geográficos, pueden ser adecuados para este tipo de tecnología. En ciertas operaciones censales —especialmente las que se realizan una sola vez—, como el ingreso de datos, la externalización puede ser una solución apropiada. Los contratistas que cuenten con el equipo y la capacitación adecuados pueden complementar al personal del censo. No obstante, la externalización también plantea problemas, entre los que cabe mencionar la superación de los obstáculos burocráticos, el manejo de los contratistas y el cumplimiento de las normas de confidencialidad.

En las últimas décadas se han logrado avances sumamente importantes en materia de cartografía censal, para la cual antes era preciso mucho trabajo sobre el terreno y mucho dibujo manual. En cambio, ahora puede realizarse mediante la teleobservación y la producción de mapas asistida por computadora. La tecnología de los sistemas de información geográfica (SIG) se está empleando cada vez más en los censos de población y de vivienda para generar mapas de enumeración y para la presentación de datos. Los sistemas de posicionamiento global son baratos y fácilmente accesibles y pueden ser utilizados por el personal de cartografía sobre el terreno para anotar los mapas topográficos y satelitales y producir mapas de excelente calidad para los enumeradores.

El software de procesamiento de datos para los censos, que antes era desarrollado y suministrado por organizaciones sin fines de lucro, está siendo reemplazado por programas que pueden obtenerse en el mercado. Sin embargo, adaptar el software de uso general a las actividades censales exige contar con una capacidad de programación considerable que, en muchos casos, no está disponible en las organizaciones censales.

La Internet, en tanto herramienta para la recopilación de datos censales todavía está en pañales, aunque en algunos países se permitió la realización de algunas tareas de enumeración a través de la web. En general, se recopilaron datos de una proporción reducida de la población, en forma experimental. Algunos de los problemas de este método son la necesidad de identificar sin lugar a dudas cada hogar; la falta de cobertura actual de muchos hogares; y el temor de que los piratas informáticos puedan poner en peligro la integridad del censo. Además, los datos recopilados deberán ser integrados a otros flujos de datos, con inclusión de los cuestionarios postales y las respuestas telefónicas. Sin embargo, como herramienta para la difusión de datos, la Internet se está convirtiendo aceleradamente en el medio más importante, y las oficinas de estadística están aumentando el

número de publicaciones electrónicas y la eficacia de los sitios web. También se están desarrollando tecnologías para el almacenamiento de datos censales, incluidos los “almacenes de datos”, donde quedarían registrados todos los datos y metadatos de los censos.

Es imposible desarrollar un conjunto único de directrices para ayudar a los encargados de la planificación de los censos a elegir las mejores tecnologías nuevas. En este sentido, la decisión depende de la magnitud del proyecto, la disponibilidad de capacidades en el plano local, la situación financiera, la experiencia previa, el tiempo disponible para los preparativos y otros factores. Los planificadores deben ser prudentes porque las soluciones propuestas deben ser funcionar desde un principio. Las nuevas tecnologías nunca pueden poner en riesgo el sistema existente de presentación de informes y, en la medida de lo posible, deben reforzarlo.

SUMMARY

Even small improvements in census technology can result in important gains in the quality and cost-effectiveness of the whole census operation. At present a number of organizations are attempting to help bring innovation to census and statistical operations. Among the concerns regarding new technology are these: how to choose appropriate technology; how to maintain the integrity of existing statistical systems; how to deal with outsourcing certain tasks; and how to maintain confidentiality of data. Some technologies, such as mobile telephony, have made person-to-person communication in the field easier, as have fax and e-mail capabilities. Barcode technology has made management of materials more efficient.

In the 2000 round of censuses, intelligent character recognition (ICR) made a breakthrough in many countries, although illegible handwritten characters and badly printed questionnaires still led to problems. In general, countries that planned carefully for the new technology and conducted pre-tests were more successful in their operations. The next step, automatic or computer-assisted coding, is also being explored, and some data, such as geographic names, may lend themselves to such coding. For some census operations, especially one-time, high-volume tasks such as data entry, outsourcing may be a good solution. Contractors with the necessary equipment and skills can supplement the census staff, but outsourcing also raises questions of overcoming bureaucratic obstacles, managing the contractor and enforcing confidentiality rules.

Census mapping has made great strides in the last few decades, from an activity requiring extensive fieldwork and manual drawing to one using remote sensing and computer-assisted map production. Geographic information system (GIS) technology is increasingly being used in population and housing censuses to generate maps for enumeration and for data presentation purposes. Global positioning systems (GPS) are cheap and available, and they can be used by cartographic field staff to annotate topographical maps and satellite photographs to produce excellent maps for enumerators.

Data-processing software for censuses, which was previously developed and provided by non-profit agencies, is being supplanted by commercially available software. However, customizing general-purpose software for census purposes requires considerable programming skills, which may not always be available in a census organization.

The Internet as a tool for census data collection is still in its infancy, although several countries did allow some Internet enumeration in their most recent censuses. Generally, such data were collected from a small portion of the population on an experimental basis. Problems with this method include the need for authentication from each household; lack of coverage of households in many countries at this stage; and the fear that hackers could compromise the integrity of the census. Moreover, data collected via the Internet would have to be integrated into other data streams, including mail-back questionnaires and telephone responses. As a tool for data dissemination, however, the Internet is quickly becoming the principal medium, and statistical offices are responding with more electronic publications and effective web sites. Technology is also under development for the storage of census data, including data “warehouses”, which would contain all the data and metadata from a census.

It is impossible to create a single set of guidelines to help census planners choose the best new technology. Choices depend on the magnitude of the project, the availability of local skills, the funding situation, prior experience, time for preparation, and other factors. Census planners need to be conservative, because their solutions must be right the first time. New technology should never endanger the continuity of existing reporting systems and if possible should reinforce it.

RÉSUMÉ

Une amélioration si légère qu'elle soit des technologies censitaires peut se traduire par un net avantage en termes de qualité et de coût de l'opération. Plusieurs organisations se sont attelées à la tâche d'introduire des innovations aux opérations censitaires et statistiques. Ces nouvelles technologies présentent néanmoins certains problèmes tels que le choix de la technologie appropriée, le maintien de l'intégrité des systèmes statistiques existants, la façon d'aborder l'externalisation de certaines activités ainsi que la préservation de la confidentialité des données. Certaines technologies, telles que la téléphonie mobile, ont facilité la communication interpersonnelle, tout comme la télécopie et le courrier électronique. Par ailleurs, la technologie des codes à barres a contribué à améliorer l'efficacité des activités de gestion des matériels utilisés dans les recensements.

Dans la série de recensements de 2000, la reconnaissance intelligente de caractères (RIC) a constitué un fait saillant dans de nombreux pays, malgré certaines difficultés au niveau de la lecture des caractères manuscrits illigibles et des formulaires dont l'impression était déficiente. D'une manière générale, les opérations des pays qui ont soigneusement l'utilisation des nouvelles technologies et ont procédé à des essais préliminaires se sont mieux déroulées. L'étape suivante est également à l'essai, à savoir la codification automatique ou assistée par ordinateur, et certaines données telles que les noms de lieu peuvent être adaptées à ce type de technologie. Pour certaines opérations censitaires, en particulier celles qui ne sont effectuées qu'une seule fois telles que la saisie des données, l'externalisation peut constituer une bonne solution. Les sous-traitants possédant le matériel et les compétences requis peuvent servir de complément au personnel chargé du recensement. Cependant, l'externalisation présente également des problèmes, en particulier en ce qui concerne les obstacles bureaucratiques, la gestion des sous-traitants et le respect des normes de confidentialité.

Des progrès considérables ont été accomplis au cours de ces dernières décennies en matière de cartographie censitaire, qui exigeait une énorme quantité de travail sur le terrain et de dessin manuel. Elle est désormais possible moyennant la téléobservation et la production de cartes assistée par ordinateur. La technologie des systèmes d'information géographique

(SIG) est de plus en plus utilisée dans les recensements démographiques et de logement pour élaborer des cartes d'énumération et la présentation des données. Les systèmes mondiaux de positionnement sont peu coûteux et d'accès facile. Ils peuvent être utilisés par le personnel de cartographie sur le terrain pour noter les cartes topographiques et satellitaires et produire des cartes d'excellente qualité à l'intention des énumérateurs.

Le logiciel de traitement des données censitaires qui jusqu'ici était mis au point et fourni par des organisations à but non lucratif est substitué par des logiciels disponibles sur le marché. L'adaptation d'un logiciel de type général aux activités censitaires requiert toutefois une énorme capacité de programmation que ne possèdent pas toujours les organisations responsables des recensements.

L'Internet, en tant qu'outil de collecte de données censitaires n'en est encore qu'à ses premiers pas bien qu'il ait servi dans certains pays à la réalisation de certaines tâches. Des données correspondant à un pourcentage réduit de la population ont été recueillies à titre expérimental. Cette méthode présente certains problèmes tels que la nécessité d'identifier chaque ménage de façon inéquivoque, le manque actuel de couverture de nombreux ménages et la crainte de la menace que représentent les pirates informatiques pour l'intégrité du recensement. Qui plus est, les données recueillies devront être intégrés à d'autres flux de données, dont les questionnaires postaux et les réponses téléphoniques. Internet devient toutefois le principal outil de diffusion des données et les bureaux de statistiques ne cessent d'accroître le nombre de publications électroniques et l'efficacité des sites internet. De même, de nouvelles technologies sont mises au point pour le stockage de données censitaires, les « mémoires », où seraient enregistrées toutes les données et méta-données des recensements.

Il est impossible d'élaborer un ensemble unique de directives permettant aux responsables de la planification des recensements de choisir les meilleures technologies de pointe. Une telle décision dépendra de l'ampleur du projet, de la disponibilité de capacités à l'échelon local, de la situation financière, de l'expérience préalable, du temps disponible pour les travaux préparatoires et d'autres facteurs. Les responsables de la planification devront faire preuve de prudence car les solutions proposées doivent pouvoir fonctionner d'emblée. Les nouvelles technologies ne doivent jamais compromettre le système actuel de présentation des rapports et doivent, dans la mesure du possible, l'étayer.

- Las viejas tecnologías nunca desaparecen por completo.
- El desarrollo de nuevas tecnologías siempre lleva más tiempo que el esperado.
- Cuando por fin se desarrollan, toman vuelo mucho más rápido de lo que se creía.

I. INTRODUCCIÓN

Es un hecho bien conocido que el arte de levantar censos se remonta a varios siglos. Desde el siglo XIX se ha tratado de aprovechar tecnologías recién desarrolladas para mejorar la eficacia y eficiencia de estas investigaciones estadísticas de gran magnitud y costo tan elevado. Los censos son actividades de gran densidad de mano de obra, para los cuales es preciso contar con mucho personal temporero. Por lo general, el componente más importante del presupuesto de un censo es el costo laboral mientras que el gasto en materia de tecnologías de la información y las comunicaciones ocupa el segundo lugar.

Una mejora incluso pequeña en las metodologías empleadas o en la eficacia del equipo utilizado puede significar una ventaja significativa en la calidad o el costo de toda la operación. Los presupuestos de los censos dependen de la estructura de costos de cada país y de la amplitud de la encuesta, pero generalmente fluctúan entre unos pocos dólares per cápita en los países de bajos costos y unos 30 dólares per cápita en los entornos altamente desarrollados. Una estimación aproximada del gasto total de la ronda actual de censos arroja una cifra de entre 30 000 y 50 000 millones de dólares. Ciertamente es una cantidad muy atractiva para quienes tratan de mejorar la relación costo-beneficio.

Entre los primeros en adaptar las tecnologías modernas a las tareas censales sobresale el nombre de Herman Hollerith, quien tomó prestadas las ideas de Joseph-Marie Jacquard, inventor de las tarjetas perforadas para controlar los telares. Hollerith se dio cuenta de que las tarjetas podrían utilizarse para el trabajo de clasificación y tabulación y de esta manera no sólo logró acelerar la publicación de los resultados del censo de 1890 de los Estados Unidos sino que inició toda una industria.

Otros innovadores menos conocidos en las actividades censales hicieron buen uso de determinados métodos y tecnologías recién descubiertos. La tecnología de la información ocupó un lugar destacado en estas innovaciones y el equipo de procesamiento de datos censales ha

evolucionado desde las máquinas que sólo facilitaban la labor de tabulación hasta convertirse en instrumentos indispensables en casi todas las fases de las operaciones censales. Las computadoras se utilizan para las tareas de planificación y apoyo a las actividades de sectorización, en la gestión de proyectos, en todas las etapas de adquisición, depuración y codificación y comunicación de datos y el análisis demográfico (Dekker, 1997). Muchos de los adelantos recientes logrados en las actividades censales han sido posibles gracias al desarrollo ininterrumpido de los equipos de procesamiento de datos y de las redes de comunicaciones que operan en los planos local, nacional y mundial. Para asegurar la continuidad, es importante que el uso de la tecnología más nueva se integre a la metodología existente y ya probada, y se desarrolle a partir de ella (Naciones Unidas, 1998).

Hay varias iniciativas importantes en curso destinadas a coordinar y orientar el proceso de innovación en las estadísticas oficiales y el levantamiento de censos. Una de ellas es el Consorcio de Estadísticas para el Desarrollo en el Siglo XXI, conocido como la iniciativa PARIS 21. Cuenta con varios centenares de miembros provenientes de las principales oficinas de estadística nacionales e internacionales, las instituciones académicas, etc. Uno de los temas que están analizando los expertos, mediante un esfuerzo mancomunado en el marco de la iniciativa, es determinar la manera de mejorar la eficacia en función de los costos de las operaciones censales. Para mayor información sobre este tema véase el sitio web <http://www.paris21.org>.

La División de Estadística de las Naciones Unidas viene impulsando desde hace mucho tiempo el uso de principios estadísticos sólidos y el intercambio de conocimientos técnicos específicos. Recientemente se ha creado un sitio web que permite acceder a información sobre las prácticas estadísticas adecuadas (<http://esa.un.org/unsd/goodprac>). En el plano regional, EUROSTAT ha organizado una serie de seminarios técnicos denominados Nuevas Técnicas y Tecnologías para las Estadísticas (NITTS) e Intercambios de Tecnología y de Conocimientos Técnicos (ETK). Las reuniones de 2001 sobre estos temas se celebraron en forma conjunta en Creta, Grecia, durante el mes de junio pasado.

También reviste interés el sitio web de EUROSTAT denominado VIROS (Instituto Virtual de Investigación sobre las Estadísticas Oficiales) (<http://europa.eu.int/en/comm/eurostat/research/viros>). En él se identifican y clasifican las áreas de investigación en las que las organizaciones participantes pueden ubicar los resultados de sus estudios y experiencias, conservando en todo momento su responsabilidad sobre ellos. EUROSTAT

funciona como ente central de coordinación, tratando de integrar los elementos individuales en un conjunto coherente, con el objetivo final de facilitar el acceso a la información sobre las actividades de investigación y sus resultados. Naturalmente, EUROSTAT se interesa en estos temas porque también se enfrenta con la necesidad de combinar muchas tradiciones estadísticas y amalgamarlas, en la medida de lo posible, con las tecnologías de integración más modernas.

Al considerar las alternativas tecnológicas que tienen ante sí, las oficinas de censos deben responder a varios interrogantes, entre los cuales cabe mencionar:

- ¿Cómo realizar una elección fundamentada de la tecnología apropiada?
- ¿Cómo mantener la integridad de los sistemas estadísticos y censales existentes?
- ¿Cómo abordar la alternativa de la externalización¹ y la gestión de las tareas externalizadas?
- ¿Cómo mantener la confidencialidad de las soluciones elegidas?

En el presente trabajo se examinarán brevemente las diversas esferas en que se han aprovechado las nuevas tecnologías en las tareas censales y se analizarán los temas recién enumerados. La respuesta definitiva a los interrogantes planteados sólo puede darla cada organización censal.

II. GESTIÓN, COMUNICACIONES, LOGÍSTICA Y GARANTÍA DE LA CALIDAD

Los censos nacionales difieren en muchos aspectos de la labor estadística cotidiana. No se realizan con la misma frecuencia que otros procesos de recopilación de datos, cuya calidad puede mejorarse gradualmente y el nivel del gasto y el número de personas necesarias son muy superiores a lo que están acostumbrados los directores de estadísticas. Por este motivo, en algunos países se crean oficinas de censos independientes del organismo nacional de estadística. Puede ser necesario contratar personal directivo habituado a manejar organizaciones temporarias de gran tamaño. Como los censos pueden considerarse grandes proyectos —en los que el tiempo es un factor decisivo— que entrañan muchas operaciones entrelazadas, el uso de software moderno de gestión de proyectos es de suma importancia.

¹ Para una definición de este y muchos otros términos empleados véase el Glosario al final del documento .

La realización de un censo exige una comunicación eficiente entre (varios) miles de personas, así como la adquisición y almacenamiento de gran diversidad de elementos, la mayoría de los cuales deben distribuirse a lo largo y a lo ancho del país, para luego volver a reunirlos.

Las nuevas tecnologías de telefonía móvil (teléfonos celulares) han facilitado la comunicación interpersonal, incluso en los países que cuentan con redes fijas extensas y confiables. No obstante, en la mayoría de los países en desarrollo no se ha logrado que todos los que trabajan en un censo puedan comunicarse por este medio, pues en algunos casos las comunicaciones con las regiones remotas siguen siendo difíciles. Es probable que los sistemas de telefonía satelital, que funcionan en cualquier lugar del mundo, llenen este vacío. En este sentido, algunos proyectos de gran envergadura, como el denominado "Iridium", no han logrado reunir un número suficiente de suscriptores iniciales. Sin embargo, en la medida en que los enormes costos de inversión ya han sido amortizados, los precios para los usuarios están bajando. Las estaciones terrestres, dotadas de las antenas necesarias, todavía son más bien voluminosas, pero totalmente portátiles. Los responsables de planificar los censos deben conocer todas las alternativas posibles de comunicaciones —con inclusión de las diferencias regionales— y hacer los arreglos necesarios.

En lo que se refiere a la transmisión de información impresa o que pueda imprimirse, los sistemas de fax están siendo reemplazados aceleradamente por el correo electrónico. Esto vale también para los censos, pero depender de este medio supone cierto grado de vulnerabilidad, por la eventual interrupción de los servicios de Internet, los problemas del analfabetismo cibernético y los virus informáticos. En tal sentido, es importante mantener siempre disponible un sistema de fax como respaldo.

Los avances en materia de software y la gran disponibilidad de computadoras personales han facilitado significativamente la gestión del transporte de bienes. En este aspecto, la tecnología de códigos de barras puede ser un elemento clave. Su empleo en lugar de números impresos tiene la ventaja de evitar los errores de transcripción y acelerar los procedimientos. Puede usarse una combinación de ambos si también se desea que las personas reconozcan los códigos con facilidad. Los directores de los censos, que no son especialistas en logística, suelen pasar por alto esta tecnología tan difundida.

Una aplicación muy común de esta tecnología es el etiquetado de todos los elementos correspondientes a un área de enumeración determinada (mapas, identificación de enumeradores, hojas de resúmenes, cajas de transporte, etc.) con un código de barras propio. En el lugar de destino de los materiales se escanearán los códigos, actualizando automáticamente

las bases de datos de los elementos remitidos. El mismo procedimiento puede utilizarse para mantener una base de datos de los elementos devueltos por quienes trabajan sobre el terreno.

El etiquetado de los cuestionarios individuales con códigos únicos también puede resultar útil, aunque los gastos generales de administración resultantes pueden ser significativos. Estos identificadores pueden servir para evitar un problema bastante frecuente, como es el hecho de que se devuelvan partidas completas de cuestionarios geocodificados erróneamente. Tanto los escáneres estándar, vendidos en el comercio, como la mayoría de los sistemas de reconocimiento inteligente de caracteres (véase la sección 3 a)) pueden leer los códigos de barras sin dificultades.

La garantía de calidad, que incluye el uso de métodos de muestreo sólidos desde un punto de vista científico, debe ser parte integrante de todas las operaciones censales. Muchos de los métodos empleados en esta esfera se basan en principios estadísticos y han sido desarrollados por innovadores en esta materia (Deming, 1986). La oficina del censo debe lograr un nivel consistente de garantía de calidad en todas sus operaciones y no puede darse el lujo de omitir las técnicas que le permitan lograrlo y verificarlo (Conferencia Internacional sobre la Calidad de las Estadísticas Oficiales, 2001).

III. ADQUISICIÓN DE DATOS

a) Reconocimiento inteligente de caracteres (RIC)

Es posible decir, sin faltar a la verdad, que la actual ronda de censos ha constituido el triunfo tecnológico del RIC. En la ronda 1985-1994 sólo aproximadamente 20% de los países que realizaron censos emplearon algún tipo de método de reconocimiento de caracteres o marcas (Dekker, 1994). La gran mayoría seguía recurriendo a la adquisición de datos mediante teclados. En la ronda actual, casi todas las oficinas de censos de las economías industrializadas de mercado, y muchas otras, aplican métodos de adquisición de imágenes por medio de escáneres, software de reconocimiento y todos los demás elementos necesarios para eliminar (parcialmente) el ingreso manual de datos.

No hay duda de que en la última década se han logrado grandes avances en materia de tecnologías de reconocimiento, pero también parece ser cierto que el ejemplo dado por los “pioneros” de los censos en esta materia ha facilitado el cambio a aquellas organizaciones que de otro modo hubieran dudado en adoptarla. El RIC parece ser un instrumento muy

promisorio para mejorar la eficiencia, pero es intrínsecamente más riesgoso que el ingreso de datos mediante el uso de teclados. Por ejemplo: si los cuestionarios están mal diseñados o su impresión es deficiente, resultan muy incómodos para el ingreso manual de datos pero pueden significar el fracaso total de una operación prevista de adquisición de datos mediante el RIC. La necesidad de pruebas exhaustivas antes de adoptar una tecnología determinada, tan evidente en los métodos censales tradicionales, lo es aún más cuando se pretende usar métodos de generación de imágenes digitales.

Sin embargo, subsiste un problema fundamental, y es que muchas veces el reconocimiento de los caracteres manuscritos es deficiente si quien los escribe no tiene conocimientos previos sobre el sistema de reconocimiento. Evidentemente, esto ocurre en los censos cuyos formularios son llenados por el propio encuestado, o aquellos en que participa un gran número de enumeradores. Para evitar este problema se puede limitar el uso del reconocimiento automático sólo a las marcas y los dígitos. No obstante, a veces en los dígitos tampoco se logra una interpretación confiable, de modo que, para salvar este inconveniente, se necesita contar con gran número de personas que ingresen los datos manualmente.

Existen informes aislados que indican que el proceso de RIC no siempre funciona tan ordenadamente como se espera. La experiencia adquirida en las pruebas finales de funcionamiento impulsó a la Oficina del Censo de los Estados Unidos a adoptar un sistema de procesamiento de dos lecturas sucesivas en lugar de una. Conforme a este sistema, los datos muestrales de los formularios largos sólo se almacenarán en la computadora durante la segunda operación de adquisición de datos (Prewitt, 2000). Este cambio de enfoque no ha modificado los plazos de procesamiento. En algunos países de Europa (como Estonia) se han reportado dificultades en el reconocimiento de caracteres alfabéticos manuscritos, lo que ha obligado a contratar personal adicional para colaborar en el proceso de reconocimiento automático. En una reunión celebrada recientemente en Bangkok (CESPAP, 2001) se mencionaron problemas de diversa gravedad en Tailandia, Filipinas, China y la Región Administrativa Especial de Macao² (para obtener información detallada

² Por ejemplo, las tasas de reconocimiento de los caracteres manuscritos pueden caer por debajo de 90%. Siempre habrá que tener en cuenta este valor en relación con el nivel de seguridad, un parámetro preestablecido que determina la "confiabilidad" que deben tener los dispositivos de reconocimiento antes de aceptar que un carácter representa un símbolo determinado. Entre los caracteres aceptados suele haber equivocaciones (los "errores"). En cambio, los caracteres rechazados pueden incluir "valores aceptados", que son caracteres que hubieran sido reconocidos correctamente con un nivel de seguridad más bajo. Los demás caracteres rechazados son "valores correctos" que siempre obligan al operador a realizar alguna operación.

sobre los problemas señalados, véanse los documentos de los países en el sitio web mencionado).

En Tailandia, los planes preliminares de establecer 15 centros regionales de RIC para el censo de abril de 2000 se cancelaron cuando se concluyó que sería necesario utilizar escáneres y software más avanzados y caros. Actualmente funciona un único complejo de RIC en Bangkok (equipado con escáneres Fujitsu 4099 y un software denominado TeleForm). También se informó de algunos problemas con los caracteres mal escritos y el mantenimiento de los escáneres.

Para el censo del 1º de mayo de 2000 de Filipinas se trabajó con cuatro centros de adquisición de datos descentralizados, empleando escáneres Kodak 3590 y un software denominado *Eyes and Hands*. En este caso, uno de los problemas más graves es que la calidad de impresión de algunos de los cuestionarios no cumple las especificaciones, por lo que el software de RIC los marca como no identificables. Otra dificultad proviene de las anotaciones manuscritas ilegibles. El número de licencias de verificación necesarias para corregir manualmente los formularios rechazados se había subestimado. Este ha sido un proceso de aprendizaje y como las experiencias han sido satisfactorias se volverá a emplear el RIC en el próximo Censo de Agricultura y Pesca.

En la Región Administrativa Especial de Macao se han obtenido buenos resultados en los ensayos piloto del censo de 2001. En el documento se incluye un cuadro interesante, obtenido a partir de una muestra de 150 000 imágenes de dígitos. Los resultados indicados en el cuadro no permiten obtener conclusiones directas que confirmen la eficacia del RIC tal como se utilizó. Podría ser útil proporcionar capacitación a los enumeradores sobre la mejor manera de escribir determinados caracteres numéricos.

Dígito	0	1	2	3	4	5	6	7	8	9	Total
Tasa de reconocimiento*	94.83	96.83	94.92	91.11	96.00	94.95	97.29	97.72	90.43	81.74	95.64
Tasa de rechazos*	5.17	3.17	5.08	8.89	4.00	5.05	2.71	2.28	9.57	18.26	4.36
Tasa de exactitud*	99.38	99.89	99.78	99.73	99.89	99.41	99.79	99.59	99.12	100.00	99.72
Tasa de errores*	0.62	0.11	0.28	0.27	0.11	0.59	0.21	0.41	0.88	0.00	0.28

* En porcentaje.

Para el censo de Indonesia del 1º de julio de 2000, el RIC se realizó en 29 centros de procesamiento en todo el país, empleando escáneres Kodak

DS 3500 y el software de reconocimiento NCS *NestorReader* incorporado a un programa propio realizado en *Visual Basic*. En el documento presentado por el país, se informa de numerosos problemas que dificultan el proceso de RIC del censo, entre los que cabe mencionar la impresión deficiente de los cuestionarios (pese a un complejo sistema de control de calidad), la escritura incorrecta de los enumeradores, el manejo inadecuado de los documentos sobre el terreno (por lo que los formularios no resultan utilizables), los problemas de mantenimiento de los escáneres y un sistema complejo de gestión de archivos. Los autores del informe merecen una calurosa felicitación por compartir estas experiencias a fin de que otros puedan aprovecharlas. Es evidente que la enorme magnitud de la operación, los disturbios en algunos puntos del país, las restricciones presupuestarias y varios problemas logísticos han incidido significativamente. Pese a las dificultades, la Oficina Central de Estadística de Indonesia confía en que la operación de adquisición de datos podrá realizarse sin tropiezos.

En el censo de Aruba de octubre de 2000 (sobre el cual no se informó en la reunión de Bangkok) se emplearon escáneres Fujitsu M3079DG y el software *Eyes and Hands*. Todos los datos de este pequeño país (aproximadamente 100 000 habitantes) se terminarán de adquirir en abril de 2001. La operación se preparó cuidadosamente y funcionó sin dificultades, con inclusión del trabajo integrado de codificación asistido por computadora. En comparación con el ingreso de datos mediante teclado no se observaron ventajas en materia de costos.

Las dificultades sobre las cuales se ha informado pueden dividirse en las relacionadas con el proceso de reconocimiento en sí mismo, y todas las demás. Si la tasa de reconocimiento está por debajo del nivel aceptable, generalmente la solución es reducir el nivel de seguridad preestablecido. Pero esto, a su vez, acarrea ciertos inconvenientes, pues las tasas de error serán mayores. También puede ocurrir que el movimiento del papel en los escáneres no sea confiable y esto puede deberse a varios motivos, como la presencia de polvo, el uso de líquido corrector en las hojas y el deterioro de formularios que puede producirse si las condiciones climáticas no son adecuadas. Se han dado casos en que, debido a estas dificultades, ha sido necesario transcribir gran número de cuestionarios, lo que también aumenta las tasas de error.

Por lo general, las oficinas de censos que no tuvieron dificultades realizaron un proceso cuidadoso y prolongado de preparación, que incluyó varias pruebas piloto. Las que se vieron obligadas a reducir el trabajo preparatorio probablemente no hayan obtenido tan buenos resultados. En este aspecto, resulta fundamental realizar un trabajo exhaustivo de gestión de la calidad, por ejemplo en el trabajo de impresión de los cuestionarios.

Si el reconocimiento del texto manuscrito se ha convertido en un instrumento más confiable, sería lógico suponer que el próximo paso será el reconocimiento de la voz. Después de todo, se trata de un método más directo de recopilación de datos. Este procedimiento tiene enormes posibilidades económicas y es objeto de amplias investigaciones. Ya están apareciendo algunas aplicaciones comerciales de esta tecnología, especialmente en el procesamiento de instrucciones verbales recibidas por teléfono, así como en la industria automotriz. No obstante, los progresos en esta esfera han sido más lentos de lo esperado. Las aplicaciones estadísticas todavía son muy escasas.

b) Codificación automática

El reconocimiento de textos orales generalmente tiene por objeto ordenar la codificación automática asociada. Es decir, la computadora lee el texto —por ejemplo el nombre de una zona geográfica— y luego selecciona el código correspondiente en un archivo o una base de datos asociados.

Estas soluciones, que en teoría permitirían completar la adquisición automática de datos y la codificación, tienen dos requisitos previos: i) que el proceso de reconocimiento sea suficientemente confiable y, ii) que los algoritmos de búsqueda permitan pasar con eficiencia de los términos reconocidos a los códigos apropiados. No es necesario que el reconocimiento de caracteres tenga un 100% de eficacia, ya que el algoritmo puede ser útil aunque los términos sean incompletos o estén parcialmente deformados.

Sin embargo, en realidad este proceso presenta varias dificultades. En primer lugar, y como se ha señalado, es preciso tener en cuenta la tasa de rechazos, que puede hacer necesaria la intervención de un número significativo de personas. En segundo lugar, cabe mencionar la dificultad de determinar automáticamente cuáles son los códigos válidos, que dependerá de las características de las variables de que se trate. En general, no es demasiado difícil codificar automáticamente los términos geográficos, excepto, quizás, los de menor nivel (como las aldeas), cuya ortografía quizás no esté estandarizada, o en el que pueden aparecer homónimos. En cuanto a la ocupación y las ramas de actividad, los problemas pueden ser mayores. Por más esfuerzos que haga el personal censal de terreno para obtener información precisa de los encuestados, en muchos casos estas variables se informarán en términos que no pueden vincularse con facilidad con las diversas clasificaciones, como la Clasificación Internacional Uniforme de Ocupaciones (CIUO), la Clasificación Industrial Internacional Uniforme

de todas las Actividades Económicas (CIIU) o la Clasificación Industrial General de Actividades Económicas de las Comunidades Europeas (NACE).

La cuestión de la codificación automática o asistida por computadora ha sido tema de numerosas investigaciones (Meyer, 1997; Dopita, 1999; y Blum, 1997). Estas actividades constituyen un desafío para quienes aplican los métodos modernos de inteligencia artificial, las redes neuronales y la lógica difusa.³ Sin embargo, por más elegantes y avanzados que sean los algoritmos de equiparación, si los informes del terreno tienen múltiples interpretaciones, son demasiado generales, o de otro modo inadecuados, el problema no es fácil de resolver. Muchos especialistas consideran que en este tipo de situaciones es difícil concebir soluciones automáticas que se aproximen, en cuanto a calidad, al juicio de un codificador experimentado. Sin embargo, es posible mejorar la eficiencia procesando los casos más sencillos mediante las computadoras y dejando el resto en manos de los codificadores.

En cuanto a la codificación de las ramas de actividad, cabe señalar que ésta puede mejorarse empleando un registro de establecimientos o empresas y sus códigos CIIU o NACE. Para los encuestados puede resultar más fácil informar el nombre de su empleador que describir la principal actividad económica de la empresa. Resulta evidente que para utilizar este enfoque es preciso contar con un registro nacional completo de las empresas.

En síntesis, es indudable que los sistemas necesarios para utilizar el RIC en los censos no pueden adquirirse directamente en el mercado sino que es necesario un diseño cuidadoso de los cuestionarios y probarlos de forma exhaustiva. Para integrar el RIC con otras operaciones asociadas, como la codificación, es preciso realizar un profundo análisis previo y contar con una estrategia perfectamente definida, cuya eficacia se debe evaluar.

c) Externalización y descentralización

Puede ser interesante considerar la posibilidad de externalizar el ingreso de los datos censales, mediante el RIC u otro método. Como se trata de una operación de gran volumen que se realiza una sola vez, pueden

³ La codificación automática puede considerarse como una forma de traducción y para ella se emplean métodos similares a los aplicados en el área muy difundida de la traducción mediante computadoras de los lenguajes naturales, cuya investigación es aún más difícil.

existir contratistas cuyo equipo y capacidad les permita ofrecer a las oficinas de censos condiciones que éstas no puedan lograr internamente. No obstante, cabe señalar que la externalización conlleva responsabilidades de contratación y supervisión que también exigen contar con los recursos necesarios. Asimismo, los problemas de confidencialidad son mucho mayores cuando los contratistas externos manejan datos personales. La garantía de calidad, de por sí un tema fundamental, se convierte en un aspecto mucho más importante si se recurre a contratistas externos (véase, por ejemplo, Whitford y Reichert, 2001). Sería interesante lograr que los contratistas trabajen en las instalaciones de la oficina de censos. Sea como fuere, el personal del contratista debería estar sujeto a normas de confidencialidad al menos tan estrictas como las que se imponen al personal temporero del censo.

Cabe señalar que aún el personal directivo, con una trayectoria impecable de gestión interna, puede encontrar dificultades para controlar el trabajo externalizado, para lo cual es preciso contar con aptitudes distintas. Entre ellas pueden mencionarse los conocimientos sobre el mercado de los servicios y las cuestiones jurídicas, las aptitudes de negociación y otras. En los censos es muy fácil que el contratista logre una posición dominante, pues por más que la oficina de censos no esté satisfecha con los servicios que presta, ya no está en condiciones de dar marcha atrás.

A veces, las normas de la administración pública constituyen un obstáculo para la externalización de determinadas tareas que sería mejor derivar a proveedores especializados, ajenos a la oficina de censos. Evidentemente, es preciso modificar esta situación, pero lo más probable es que las reformas necesarias deban adoptarse en un nivel del gobierno distinto al que pertenece el encargado de supervisar los servicios nacionales de estadística.

La descentralización del proceso de adquisición de datos permitiría a la organización del censo mantener el control sobre las actividades, y al mismo tiempo aprovechar las ventajas de derivar el trabajo a los centros regionales. Los problemas son de algún modo comparables a los de la externalización, pero más fáciles de manejar. Muchos aspectos dependen de la situación del país de que se trate: la magnitud de la tarea, las condiciones del mercado de trabajo, la eficiencia de las comunicaciones y de los medios de transporte, entre otros. La realización de mayor número de actividades fuera de la capital también puede reportar ventajas de carácter social y en el área de las relaciones públicas. Es imposible formular directrices generales en esta esfera.

IV. LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICA (SIG), LA TELEOBSERVACIÓN Y LOS SISTEMAS DE POSICIONAMIENTO MUNDIAL

En el documento titulado “Identifying and resolving problems of census mapping”, presentado en esta misma reunión, puede encontrarse un análisis más exhaustivo de estos temas. Como las nuevas tecnologías de cartografía son una parte esencial de las innovaciones en materia de actividades censales, a continuación se formulan algunas breves observaciones.

En las últimas décadas la tecnología de la cartografía ha avanzado considerablemente. Ha pasado de ser una actividad que dependía de la exploración sobre el terreno y del dibujo manual, a otra que emplea la teleobservación y la gestión cartográfica asistida por computadora.

Si bien antes de las tecnologías satelitales se había empleado la aerofotografía para la cartografía censal (especialmente en las zonas urbanas densamente pobladas), aquellas constituyen una solución más eficaz en función de los costos en el área de la teleobservación. La resolución de las fotografías satelitales disponibles en el mercado actual es muy superior a la necesaria para identificar cada edificio. La existencia de estas fotografías reduce de forma considerable la necesidad de la inspección sobre el terreno aunque no la elimina por completo.

Los sistemas de posicionamiento mundial de mano, baratos y ampliamente difundidos, que a su vez dependen de la tecnología satelital, facilitan considerablemente las tareas de inspección mencionadas. Los mapas topográficos y las fotos satelitales constituyen el punto de partida de la cartografía censal sobre el terreno. El personal de cartografía, equipado con mapas, fotografías y sistemas de posicionamiento, puede completar y anotar los mapas, que constituyen un excelente material de orientación para los enumeradores.

Actualmente, los mapas se producen, almacenan y actualizan utilizando sistemas computarizados especializados y software comercial. Los elementos esenciales de las fotografías satelitales o los mapas impresos pueden digitalizarse por rastreo manual en mesas digitalizadoras. Una vez terminados, los mapas pueden imprimirse o reimprimirse a voluntad. Las imágenes vectoriales se almacenan en archivos de computadoras sin peligro de que se deterioren con el tiempo.

En este contexto, cabe señalar que existe una tendencia cada vez más pronunciada de los organismos nacionales de estadística a definir zonas básicas de información estadística, sin tener en cuenta la organización territorial administrativa (Jacob, 1999). A veces, se emplean cuadrículas

para este fin.⁴ Las zonas de información deberían ser lo suficientemente amplias como para mantener la confidencialidad de las respuestas individuales, pero lo suficientemente pequeñas como para reagruparlas y de hacerlas coincidir con las unidades mínimas de ordenación territorial. Con este método se eliminan algunos de los problemas que entraña mantener series temporales en un contexto siempre cambiantes de límites administrativos.

El valor de la información censal es mucho mayor si se combina con los mapas básicos subyacentes, a fin de que los usuarios puedan generar mapas temáticos según sus necesidades. Actualmente, varias oficinas censales comercializan productos integrados (generalmente en CD-ROM) que ofrecen esta posibilidad. Otras oficinas consideran que tales servicios trascienden las atribuciones de las agencias nacionales de estadística y se limitan a proporcionar datos censales agregados a las editoriales comerciales. Esto no debe confundirse con la externalización, ya que la responsabilidad del producto final corresponde a la contraparte. La única responsabilidad de la oficina de censos es suministrar datos confiables, respetando las normas de control de la confidencialidad de las estadísticas.

Muchas oficinas de estadística mantienen uno o varios sistemas geográficos de información para su propio uso. Al mismo tiempo, es un hecho generalmente aceptado que la función más importante de los estadígrafos es proporcionar datos de la mejor calidad posible a los usuarios. En muchos casos, es mejor dejar en manos de otros la tarea de integrar la información proveniente de varias fuentes a los sistemas de información geográfica complejos. Esta afirmación es válida sobre todo si estos sistemas prestan servicios a una comunidad de usuarios especializados, como los urbanistas o los especialistas en medio ambiente.

Los mapas electrónicos se han convertido en instrumentos indispensables y eficaces en función de los costos para una amplia gama de operaciones censales y estadísticas.

⁴ Desde 1968, la Oficina Federal de Estadística de Suiza ha utilizado una cuadrícula muy estrecha, de 100 metros de lado, principalmente para las estadísticas ambientales y agrícolas.

V. PROCESAMIENTO Y ALMACENAMIENTO DE DATOS

a) Software de procesamiento de datos censales

En muchos países, especialmente los del mundo en desarrollo, se ha recurrido desde hace tiempo al software de dominio público para satisfacer las necesidades de procesamiento de los datos censales. Este software fue desarrollado y mantenido por organizaciones sin fines de lucro, apoyadas generalmente por subsidios de los donantes nacionales o internacionales.

Parecería que, en general, las actividades de desarrollo de este tipo de software han sido menores en los últimos tiempos que cuando se realizaron las rondas censales anteriores. Esto puede explicarse, en parte, por la capacidad cada vez mayor del software disponible en el mercado. Quizás también se deba a cierta fatiga de los donantes. En general, éstos prefieren pensar en términos de proyectos con un principio y un fin bien definidos. El desarrollo y el mantenimiento de los sistemas de software es una tarea que jamás termina pues la evolución de los entornos de hardware y de software obliga a realizar actividades de apoyo y actualización permanentes, y éstas pueden ser considerables.

Como las actividades de desarrollo de nuevos sistemas de dominio público de procesamiento de datos censales o de encuestas (o de actualización de los existentes) son escasas, aquellos están quedando algo obsoletos. En algunos, se basan total o parcialmente en el sistema operativo DOS. Aun cuando este tipo de software siga siendo tan eficaz como antes y esté en perfectas condiciones de realizar las operaciones necesarias, la interfaz DOS es poco conocida por la nueva generación de usuarios. Éstos, a su vez, pueden tener dificultades para convencer a sus supervisores o a sus pares de que es preferible trabajar en un entorno aparentemente desactualizado. El uso de herramientas modernas favorece la reputación de los especialistas en procesamiento de datos. Parecería que esto ha llevado a utilizar cada vez más programas alternativos, como los sistemas comerciales de software estadístico (SAS, SPSS, entre otros) y aplicaciones de generación de bases de datos (Microsoft Access).

Algunos anuncios recientes han mejorado el panorama del software no comercial. La Oficina del Censo de los Estados Unidos, a través del Centro Internacional de Programas, está ofreciendo módulos adicionales de su sistema CPro Census and Survey Processing, cuyo desarrollo está realizando en cooperación con Macro International y Serpro S.A.

(<http://www.census.gov/ipc/www/cspro>). El Centro Latinoamericano y Caribeño de Demografía (CELADE), División de Población de la Comisión Económica para América Latina y el Caribe (CEPAL), continúa con su labor de desarrollo de versiones más avanzadas del sistema de bases de datos estadísticos, denominado REDATAM (recuperación de datos para áreas pequeñas por microcomputador) (<http://www.cepal.cl/celade-eng>).

El desarrollo de aplicaciones de procesamiento de datos censales con software cuyo objeto original era otro puede definirse como la adaptación de esas aplicaciones a los censos. Para ello es preciso contar con capacidades de programación que no siempre se consiguen con facilidad. Parece casi inevitable utilizar en alguna medida lenguajes de programación modernos orientados a objetos. Tampoco existen organizaciones que cuenten con programas de capacitación en un tema tan especializado (como el desarrollo de aplicaciones censales a partir de software de uso general). En consecuencia, las oficinas de censos han recurrido a contratistas externos que no tenían un conocimiento suficientemente amplio de los aspectos estadísticos correspondientes. En este sentido, la situación actual con respecto al procesamiento de los datos censales es más compleja que la existente en las rondas de censos precedentes.

Por otra parte, si la oficina de censos dispone desde un principio de capacidades informáticas básicas suficientes, el personal encargado del procesamiento de los datos censales puede haber tenido mayor contacto con software moderno de uso general. Esto puede tener efectos indirectos en otros trabajos de desarrollo estadístico, en las carreras de aquellos o (cabe esperar) en ambos.

No pueden subestimarse las dificultades de adaptar el software de uso general al procesamiento de datos censales. Puede resultar mucho más complejo que emplear software especializado en este tema. La externalización de esta tarea puede llegar a multiplicar los problemas. Habrá que asegurar que los contratistas a quienes se confía el desarrollo de sistemas de procesamiento de datos censales tengan antecedentes comprobados en esta área. Sea como fuere, la oficina de censos deberá contar con los conocimientos especializados necesarios para emprender la tarea de contratación y supervisión de las actividades.

El tema más amplio del control de la confidencialidad de las estadísticas (con inclusión del uso de software de supresión de celdas) que deben ejercer todas las oficinas de censos para proteger la confidencialidad de las respuestas individuales se analizará brevemente en la sección 7 a).

b) Almacenamiento de datos

Los datos censales se almacenaban simplemente como archivos planos. Una preocupación importante era asegurar que los datos y la metainformación pudieran conservarse en buenas condiciones durante largo tiempo. Esto tenía por objeto garantizar que en una etapa posterior sería posible realizar otros análisis computarizados, por ejemplo en el censo siguiente. Actualmente, los organismos de estadística tienen cada vez mayor conciencia de que los datos obtenidos en varias recopilaciones pueden tener mucho mayor valor agregado si se conservan (con los metadatos asociados) en una estructura común de almacenamiento, a veces denominada "almacén de datos". Si bien este término tan de moda puede desaparecer tan rápido como se impuso, el principio subyacente sigue siendo válido. Se ha examinado el modelo de almacenamiento relacional para depositar información estadística, pero no siempre resulta satisfactorio. En la sección 7 c), dedicada a los ficheros estructurados, se examinará este tema con más detalle.

VI. EL USO DE LA INTERNET

a) La Internet para la recopilación de datos

Si bien el correo electrónico ha sido bastante común desde fines de los años ochenta, el acceso a los contenidos de la Internet a velocidades de transmisión aceptables no era frecuente durante la ronda de censos anterior. Sin embargo, los

En el 53° período de sesiones del Instituto Internacional de Estadística (ISI), que se celebrará este mes en Seúl, se realizará una Reunión de Presentación de Trabajos sobre éste y otros temas conexos. Se trata de la Reunión 17, sobre "Internet Data and Innovative Collection", convocada por Warren Mitofsky de los Estados Unidos de América.

problemas con los cuestionarios impresos se habían hecho evidentes mucho antes. En muchos países, la tasa de respuestas a los cuestionarios enviados por correo están disminuyendo como consecuencia del cansancio de los encuestados y, quizás, de su menor sentido cívico. Si bien los enumeradores todavía visitan personalmente los hogares, la posibilidad de encontrar a los encuestados en su casa durante las horas de trabajo han disminuido, habida cuenta del estilo de vida actual y la disminución del número de miembros de aquéllos.

Las oficinas censales han propuesto o utilizado diversas medidas para solucionar estas dificultades. Entre ellas cabe mencionar la realización de campañas de información más complejas y los esfuerzos por mejorar la cooperación de la sociedad civil. Así, los enumeradores trabajan los fines de semana y en horario vespertino, se comunican con los encuestados por teléfono y realizan un muestreo inicial del conjunto de personas que no responderían a la encuesta (con lo que, en cierto modo, se renuncia a la cobertura total). Si bien se ha informado de algunos éxitos, el esfuerzo y los gastos necesarios para lograr una tasa de respuestas aceptable son ahora considerablemente mayores que antes.

Así, resulta lógico que se haya centrado la atención en la Internet como portal de acceso a un número cada vez mayor de hogares. Empleando tecnologías de alimentación activa se podría enviar a todos los hogares conectados a la Internet un cuestionario electrónico dotado de una identificación exclusiva que, en la medida de lo posible, ya incluyera los datos básicos obtenidos en el registro civil. Los encuestados corregirían y completarían la información y luego reenviarían el cuestionario por el mismo medio a la oficina del censo, la que recibiría un registro electrónico que evitaría buena parte de la labor de ingreso de datos.

La recopilación de datos electrónicos de los establecimientos (como las empresas y los organismos del gobierno o la administración pública) se ha convertido en un procedimiento bastante común. Empleando el mismo método para los hogares y las personas se podrían utilizar métodos de entrevistas asistidas por computadora (conocidas por su sigla en inglés, CASI) (Figueiredo, 1999; Keller, 1999), que pueden resultar muy útiles para los encuestados y evitar los errores.

Lamentablemente, todavía existen varios problemas que impiden la recopilación de datos electrónicos de los hogares:

- La cobertura es incompleta: si bien el número de hogares con acceso a la Internet está aumentando rápidamente en casi en todo el mundo, sólo en unos pocos países la tasa de conexión ha superado el 50%;
- Los problemas de sesgo: el acceso a Internet es más común en los hogares de mayores ingresos integrados por personas jóvenes, por lo que la generalización de este método de recopilación de datos puede dar lugar a un patrón de respuestas sesgado;
- La carencia de un sistema de direcciones estructurado: en comparación con el correo o la red telefónica, el sistema de direcciones de Internet está mucho menos regulado, hecho que refleja el origen de este

sistema. En general, los usuarios inventan sus propias direcciones, las modifican con frecuencia y pueden tener una o varias. Reunir las direcciones de correo electrónico vigentes en un momento determinado puede ser una tarea ímproba y además sería prácticamente imposible mantener un registro de estas características con cierto grado de confiabilidad. Por el momento, estos factores impiden el uso de las tecnologías de alimentación activa para la realización de los censos;

- La atracción para los piratas informáticos: es muy probable que el uso de la Internet para la realización de encuestas atrajera a los piratas informáticos, para quienes sería un desafío ser enumerados dos veces, utilizar la identificación de otra persona o realizar actos aún más graves. Resulta comprensible que los funcionarios de los censos no tengan mayor interés en enfrentar estos desafíos.

Pese a estas dificultades, varias oficinas censales como las de Suiza (véase gráfico 1), los Estados Unidos y Singapur han permitido las respuestas electrónicas durante la actual ronda de censos (<http://www.statistik.admin.ch>; Haug, 2000; Prewitt, 2000, CESPAP, 2001). En

Gráfico 1
RECOPIACIÓN DE DATOS CENSALES EN SUIZA
A TRAVÉS DE LA INTERNET
(Versión demostrativa, vista parcial)

e-census Login

Vous pouvez entrer ici vos codes personnels de sécurité.

Lorsque vous aurez entré toutes les informations, cliquez sur la touche WEITER». Vos indications seront alors vérifiées et l'écran affichera l'emplacement de la première question. Cette opération peut prendre plusieurs secondes.

Identificateur: MUSTERPIERRE

mot de passe: *****

Les indications pour les champs ci-dessus figurent dans le questionnaire de ménage brun. Elles doivent être saisies telles qu'elles apparaissent sur le questionnaire.

Important: ne transmettre en aucun cas votre identificateur et votre mot de passe à des tiers.

Lorsque vous aurez entré toutes les informations, cliquez sur la touche WEITER». Vos indications seront alors vérifiées et l'écran affichera l'emplacement de la première question. En raison de la procédure complexe de sécurité, cette opération peut prendre plusieurs secondes.

Office fédéral de la statistique, Recensement de la population 2000, 2010 Neuchâtel

este caso no se emplearon técnicas de alimentación activa. Más bien, se pedía a los encuestados que tomaran la iniciativa, bajando los formularios censales o completándolos mientras estaban conectados con el sitio web de la oficina de censos. Para evitar que este medio se utilice indebidamente, es fundamental que cada hogar pueda autenticar la respuesta. Esto puede entrañar la certificación mediante un código de identificación exclusivo, al que los demás no tuvieran acceso. Así, el código debe enviarse al hogar, por ejemplo mediante un sistema de entrega en mano. El envío seguro y confiable de los códigos de validación por medios electrónicos también es un problema difícil de resolver en la situación actual de la tecnología. Las respuestas electrónicas ingresadas al navegador deben encriptarse, pues si carecen de este tipo de protección podrían ser interceptadas.

En el censo de los Estados Unidos las respuestas por Internet se limitaron al formulario breve, y no se hicieron mayores esfuerzos por publicitar o recomendar este método. Parece que en los tres países citados, la demostración de que la organización censal posee la tecnología más actualizada fue un factor que contribuyó a abrir el canal de la Internet.

Como parece poco probable que en el corto plazo puedan eliminarse los cuestionarios impresos, la recopilación de datos a través de la Internet exige una integración cuidadosa de los dos flujos de datos, o tres en el caso de Singapur, donde otra alternativa fueron las respuestas telefónicas (CATI, véase el glosario).

Varias oficinas de censos, como la Oficina Nacional de Estadística del Reino Unido, han informado de su decisión de no utilizar todavía la Internet para la recopilación de datos luego de evaluar los riesgos que supone navegar por aguas aún inexploradas. La Oficina de Estadística de Canadá (Statistics Canada) estuvo realizando una "prueba de Internet" en dos zonas geográficas bien definidas durante el censo de 15 de mayo de 2001.

En conclusión: subsisten múltiples problemas, de diverso carácter, que hasta ahora han impedido el uso generalizado de cuestionarios electrónicos con fines censales. La Internet aún debe crecer y es preciso desarrollar y probar métodos adecuados de recopilación de datos censales por este medio. Según las expectativas, el tema de las respuestas electrónicas habrá evolucionado significativamente para la próxima ronda de censos.

b) Uso de la Internet para la difusión de datos

La tecnología de difusión de información estadística está sufriendo un cambio profundo. Por cierto, las publicaciones impresas no han desaparecido y siguen desempeñando un papel importante, por ejemplo,

En el 53º período de sesiones del Instituto Internacional de Estadística (ISI), que se celebrará este mes en Seúl, se realizará una Reunión de Presentación de Trabajos sobre este y otros temas conexos. Se trata de la Reunión 17, sobre "Internet and Innovative Data Dissemination", convocada por Heli Jeskanen-Sundström de Finlandia.

como registro permanente y accesible en forma continua, que facilita las búsquedas. Sin embargo, las consultas en línea de los sitios web de estadística, que brindan información, gratuita o no, se están convirtiendo en el principal instrumento de difusión de la información. Esto se realiza a través de la Internet ya que los tableros de anuncios electrónicos, a los que debe accederse mediante las comunicaciones punto a punto con el proveedor de información, no resultan tan cómodos para el usuario.

Para las oficinas de estadística, esto constituye un desafío enorme. Acostumbradas a la tranquilidad relativa que suponía preparar cuidadosamente una publicación y luego esperar a que saliera impresa, ahora deben cumplir un calendario estricto de difusión electrónica. Los usuarios siempre quieren acceder a los datos lo más pronto posible pero luego se quejan cuando éstos deben ser revisados o, peor aún, cuando resulta que tienen algún error.

En estas condiciones, la elaboración de una estrategia de difusión no se ha simplificado en absoluto. La comunidad de usuarios espera, con razón, que las oficinas de estadística aprovechen en forma integral los nuevos sistemas de difusión. No obstante, sigue habiendo una demanda significativa de publicaciones impresas. Esto puede ocurrir en situaciones en que los fondos o las capacidades técnicas son escasos. Las oficinas de estadística no sólo deben formular una estrategia sino que la deben revisar periódicamente. Cuando los costos así lo requieren, es preciso ajustar los puntos de distribución teniendo en cuenta la información sobre su uso. La aplicación de mecanismos de recuperación de costos pueden contribuir a mejorar la situación.

Como ocurre con las publicaciones impresas, la información electrónica puede tener una calidad cognoscitiva muy heterogénea, incluso más que aquéllas. Además, dada la velocidad del desarrollo tecnológico, elegir la mejor interfaz posible es un objetivo muy difícil de precisar. En la Nomenclatura sobre Investigaciones Estadísticas (NORIS) de EUROSTAT

(<http://europa.eu.int/en/comm/eurostat/research/viros>) se identifican los ejemplos siguientes de investigación en esta esfera:

- Contribución a las actividades de normalización relacionadas con la Internet a fin de que puedan tenerse en cuenta las necesidades estadísticas;
- Empleo de aplicaciones de alta densidad de sistemas de banda ancha: encuestas estadísticas; difusión por audio y vídeo;
- Uso de agentes inteligentes, como los robots de conocimientos (*knowbots*), para el intercambio de información;
- Perfeccionamiento de las interfaces entre las personas y las máquinas, con inclusión del uso de sistemas de realidad virtual;
- Aplicación de tecnologías para mejorar la visualización de la información estadística vinculada con los aspectos geográficos.

Las características del sitio web de una organización de estadística se están convirtiendo en un elemento cada vez más importante. Hoy día, la evaluación de las organizaciones estadísticas o censales no sólo se hace sobre la base de la calidad y la oportunidad de la información impresa sino, también —y quizás en mayor medida— teniendo en cuenta la efectividad de su presencia en la Internet.

Es preciso adoptar las medidas adecuadas. El desarrollo y mantenimiento de los sitios web debe estar en manos de profesionales. Habría que hacer todo lo posible para implantar un sistema permanente de monitoreo de la satisfacción de los usuarios y de los hábitos de navegación de los visitantes, a fin de facilitar el acceso a los rubros que despiertan más interés, registrar los signos de confusión que pudieran tener los usuarios y asegurar el perfeccionamiento permanente del sitio en general. Si éste posibilita el acceso dinámico a ciertas bases de datos, las aplicaciones correspondientes deberían ser razonablemente confiables y haber logrado madurez suficiente (CESPAP, 2001). Un servicio de alta tecnología que sólo produzca desaliento a gran número de usuarios no beneficia a nadie.

La necesidad de disponer de una amplia gama de capacidades actualizadas sobre las tecnologías de la información y las comunicaciones y de conocimientos prácticos sobre la Internet, en un entorno en que la demanda de este tipo de aptitudes es muy grande, constituye una carga muy pesada para muchos organismos nacionales de estadística. Una solución puede ser la externalización, pero como la difusión de información es una actividad central de estos organismos, no puede considerarse automáticamente como la mejor alternativa.

De paso, cabe mencionar que la Internet ofrece excelentes posibilidades de difundir normas y directrices internacionales para la labor estadística y acceder a ellas. Un ejemplo de ello es el servidor de clasificaciones RAMON, desarrollado por EUROSTAT (<http://europa.eu.int/comm/eurostat/ramon>).

VII. OTROS TEMAS VINCULADOS CON LA DIVULGACIÓN DE LOS DATOS ESTADÍSTICOS

a) Control de la confidencialidad de las estadísticas

A medida que aumenta el volumen de información estadística fácilmente accesible, se hace cada vez más urgente mejorar los sistemas de protección de la información personal proporcionada por las personas o las instituciones, empleando técnicas denominadas de control de la confidencialidad de las estadísticas. Esta situación coloca a las oficinas de estadística en una situación de desventaja ya que deben proporcionar mayor volumen de información con más rapidez, mientras que los usuarios malintencionados, que intentan obtener información confidencial, disponen de herramientas analíticas informáticas cada vez más poderosas y el tiempo juega a su favor. Actualmente, resulta poco práctico inspeccionar visualmente cada cuadro o cubo de datos (véase *infra*) para determinar los riesgos potenciales. No obstante, ya existen herramientas de tamizado automático que ayudan en esta tarea (Willenborg y de Waal, 2001; Giessing, 1999). Éstas pueden eliminar, combinar o disimular de alguna manera los valores de las celdas que presentan riesgos potenciales.

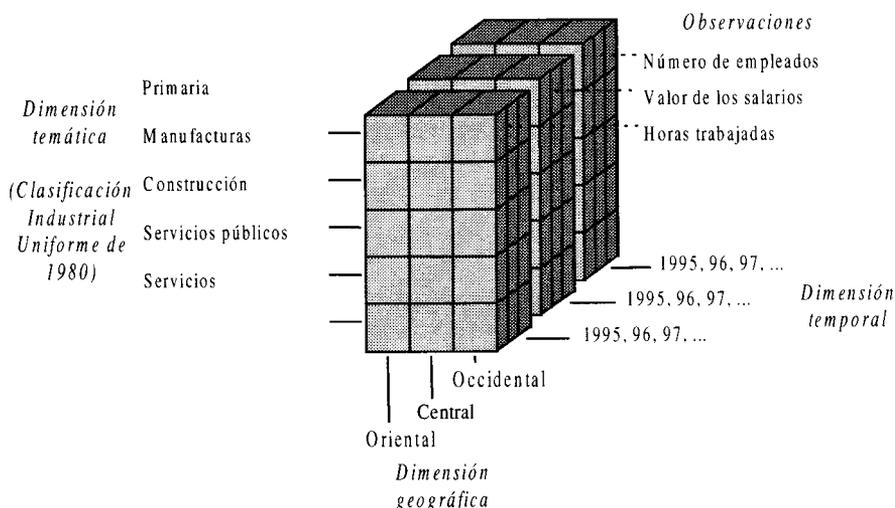
b) Medios físicos de gran capacidad

La difusión de la información en medios de almacenamiento de gran capacidad, en los que los datos no pueden modificarse, sigue siendo un instrumento importante de divulgación, especialmente cuando el volumen de los datos es muy grande y la fecha de publicación es más flexible, como ocurre con la mayor parte de la información censal. Hoy día, es muy común que esta información se publique en CD-ROM, y ya han aparecido productos en DVD de mucha más capacidad (Oficina del Censo de los Estados Unidos, 2001). La estructura de los datos en CD-ROM y la de los que aparecen en un sitio web pueden tener varias características comunes, como la navegación mediante enlaces de hipertexto. Esto facilita el desarrollo de aplicaciones paralelas.

c) Archivos estructurados: el “almacén” de datos estadísticos

Como se ha señalado en la sección 5.2, el almacenamiento de datos censales en una estructura de “almacén” facilita su uso conjunto con otra información estadística almacenada de la misma manera. Estrictamente hablando, éste no es un tema relacionado con los censos, pues se refiere a un tema más amplio de gestión de la información estadística. Un “almacén” puede estar constituido por un cierto número de cubos de datos, espacios de n dimensiones, en el que una dimensión está constituida por observaciones y las otras son dimensiones de selección. En un ejemplo sencillo de un cubo censal, las observaciones podrían ser el número total de hombres y mujeres y las dimensiones de selección los grupos de edad, el lugar de residencia, el origen étnico, la ocupación, y así sucesivamente. Para un ejemplo de cuatro dimensiones véase el diagrama del gráfico 2 (que incluye tres subestructuras tridimensionales) correspondiente a estadísticas de empresas.

Gráfico 2
CUBO DE DATOS DE CUATRO DIMENSIONES
(Basset, 1996)



Para acceder a los cubos es preciso contar con una superestructura de menús de selección por niveles jerárquicos y una combinación lógica de palabras clave. También es necesario tener acceso a los metadatos, almacenados evitando las posibles redundancias. Además de los cubos de

datos, el “almacén” también debería permitir el empleo otros formatos de almacenamiento.

Si bien existen varias aplicaciones prácticas a las que se puede acceder a través de los sitios web de varios organismos nacionales de estadística, éste es un tema sobre el que se sigue trabajando. Los “almacenes” de datos no se limitan sólo a las estadísticas, y este asunto es demasiado amplio para abordarlo en el presente documento. Para mayor información, véase, por ejemplo (Kambayashi, 2000) o búsquese en la Internet.

Llevando este concepto a otro nivel, se podría establecer que, una vez terminado el censo o la encuesta, la información recopilada debe almacenarse inmediatamente en el “almacén” de datos y que las publicaciones periódicas o únicas sólo se produzcan recuperando los datos de este sistema central de almacenamiento.

Este concepto se ilustra en el ejemplo del gráfico 3 (Keller, 2000). BaseLine es el producto final de los registros y encuestas. Contiene todos los datos obtenidos de las fuentes primarias y secundarias. MicroBase contiene los datos obtenidos en el proceso de depuración, imputación, traducción y microintegración. La base de datos de productos estadísticos agregados, denominada StatBase, contiene los resultados obtenidos tras la estimación de las poblaciones o subpoblaciones de unidades estadísticas. Se afirma que esta base de datos contiene *todos los datos publicables* producidos por la Oficina de Estadística de los Países Bajos. El “almacén” de datos para publicación, denominado StatLine, puede considerarse como un conjunto de vistas de StatBase. Presenta toda la producción de la Oficina, estructurada en forma de un conjunto de cuadros multidimensionales. StatLine se distribuye en CD-ROM o puede consultarse en la Internet.

Varias oficinas nacionales de estadística han logrado desarrollos importantes en este tema, como la Oficina de Estadística de Canadá (CANSIM II), la Oficina de Estadística de Suecia (PC-Axis) y la Oficina de Estadística de los Países Bajos. Puede obtenerse en la Internet una versión de prueba de la serie de programas Stat (el nombre del paquete completo es StatSuite) (<http://www.cbs.nl>); del sitio web de la Oficina de Estadística de Suecia puede bajarse software de recuperación de PC-Axis (<http://www.scb.se>). Otros desarrolladores de software también pueden estar dispuestos a facilitar versiones de prueba de sus programas, si se les solicita.

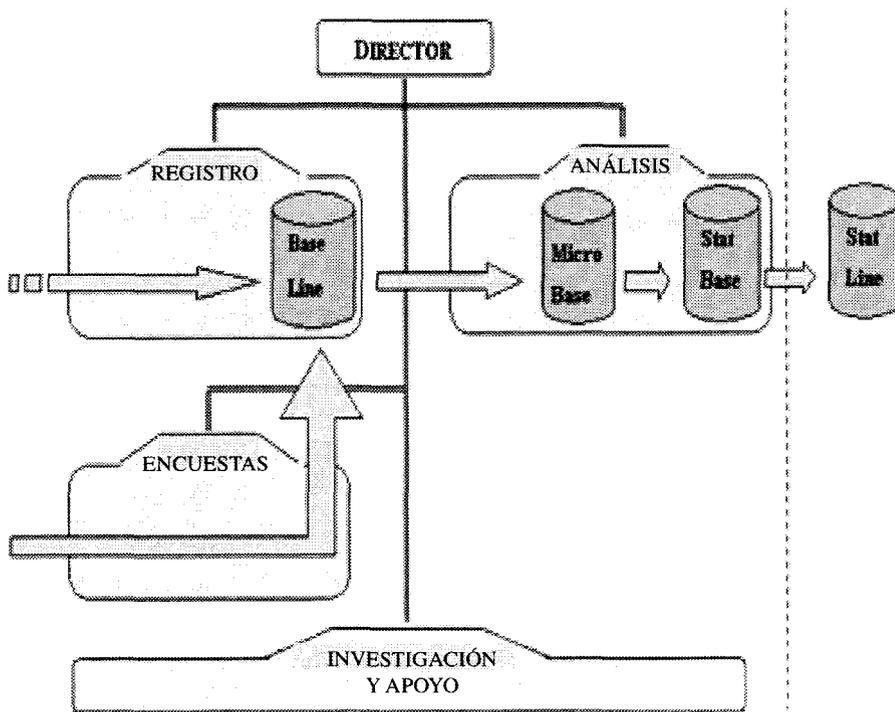
En cuanto a la aplicabilidad de los cubos de datos, parecería que el problema más importante no es tanto su almacenamiento y recuperación sino, más bien, el diseño lógico de estos repositorios de información. No es tarea sencilla formular un conjunto de cubos que cumplan con los

requisitos necesarios: i) dar cabida con facilidad a los resultados de la tarea de recopilación de datos estadísticos; ii) satisfacer las necesidades de una amplia gama de usuarios; y iii) cumplir estrictamente los requisitos de confidencialidad.

Para quienes proyectan el desarrollo de un producto de divulgación de datos, o la modificación de uno existente, no puede hacerse hincapié suficientemente en la necesidad de realizar una amplia tarea de desarrollo de prototipos y distribuir versiones “beta” a usuarios reales, dotados de criterio crítico. Esto también se ha destacado recientemente, y con argumentos muy convincentes, en el Taller sobre las Tecnologías para el Análisis, el Almacenamiento y la Divulgación de Datos sobre Población organizado por la Comisión Económica y Social para Asia y el Pacífico. Se recomienda enfáticamente la lectura del informe de este Taller y de varios otros estudios presentados en él (CESPAP, 2001).

Gráfico 3

ORGANIGRAMA DE LA DIVISIÓN DE ESTADÍSTICAS EMPRESARIALES Y SOCIALES DE LA OFICINA DE ESTADÍSTICA DE LOS PAÍSES BAJOS



VIII. CÓMO ELEGIR LA TECNOLOGÍA APROPIADA

Después de todo lo dicho, hubiera sido útil delinear directrices claras para los encargados de la planificación de los censos sobre la forma de elegir, con conocimiento de causa, la tecnología más apropiada y sobre otros temas como la externalización y los riesgos asociados con ella. Lamentablemente esto resulta imposible, pues las condiciones y las consideraciones varían ampliamente no sólo de un país al otro, sino también, y cada vez más, con el tiempo.

No existe un conjunto único preferible de tecnologías para las operaciones censales. La mejor alternativa depende de la magnitud del proyecto, las capacidades disponibles en el país, la situación financiera, la experiencia anterior, el tiempo de preparación disponible y muchos otros factores. En la actual ronda de censos se observa que la variedad de métodos y técnicas utilizados es sorprendentemente amplia.

Para los responsables, es imposible elegir alternativas fundadas si no disponen de la información necesaria. Quienes planifican los censos deben familiarizarse con los últimos adelantos en esta materia, en los planos nacional e internacional. En lo posible, deberían viajar a países similares, que hayan utilizado recientemente métodos y tecnologías que puedan resultar de interés. Sus superiores deben reconocer la necesidad de estas investigaciones y asignar los recursos necesarios.

Para definir los parámetros que se emplearán en el nuevo censo quizás sea conveniente analizar en primer lugar el censo anterior. ¿Qué cosas funcionaron adecuadamente y cuáles pueden mejorarse? Si un enfoque resultó satisfactorio en esa oportunidad, los argumentos para reemplazarlo por otro deben ser muy sólidos.

En toda decisión incide un aspecto financiero. Si los costos del censo pueden reducirse significativamente, manteniendo o incluso mejorando la calidad, es indudable que la cuestión merece un análisis detenido.

La externalización no es de ningún modo la panacea que sostienen algunos. En tal sentido, los éxitos se equiparan con los fracasos. En este tema, como en otros, no puede obviarse un análisis racional de los hechos, una negociación cuidadosa —que asegure que las posibilidades de que surjan malentendidos sean mínimas— y un programa permanente de garantía de la calidad (Whitford, y Reichert, 2001).

El último aspecto, y el más importante, debe ser el siguiente: ¿qué efectos tienen las alternativas disponibles sobre la calidad de los servicios prestados a los usuarios de la información? Las oficinas de estadística y las organizaciones censales existen gracias a los servicios que prestan a

terceros. Deben procurar, permanentemente, mejorar la información, tanto en lo que se refiere a su puntualidad, la calidad de los datos, la facilidad de acceso, la compleción y la pertinencia. Cualquier posibilidad de mejorar alguno de estos aspectos merece ser analizada.

IX. OTRAS CONSIDERACIONES

No existen demasiadas dudas de que los cambios permanentes del entorno tecnológico habrán de tener un efecto aún más pronunciado sobre los métodos censales (quizás moderado por las normas jurídicas y los problemas relativos a la confidencialidad).

Ya se ha logrado identificar a cada persona a través de determinadas características fisiológicas, como las huellas digitales o los patrones del iris, los identificadores faciales o las secuencias de la frecuencia vocal (huellas de voz) (Jain, 1999). Estas tecnologías tienen un futuro promisorio, debido a sus aplicaciones, demasiado numerosas como para enumerarlas en el presente documento (transacciones en los cajeros automáticos, control de acceso a los edificios, autorizaciones de pago, entre otros). Pueden contribuir a evitar las complicaciones que supone la necesidad de portar permanentemente documentos de identificación y tarjetas de crédito y, quizás, de recordar una serie de códigos de identificación personal.

La identificación biométrica, combinada con el acceso a una base de datos (posiblemente a través de las comunicaciones inalámbricas), puede evitar a los estadísticos la necesidad de solicitar a los encuestados una y otra vez cierta información que no se modifica (la fecha de nacimiento, el sexo, el origen étnico, el lugar de residencia en el censo anterior, entre otros). En un sentido más amplio, facilitaría técnicamente el establecimiento y mantenimiento de registros civiles electrónicos más completos y actualizados que los existentes.

Como paso intermedio podría concebirse una tarjeta inteligente personal multipropósito, de la que se podría acopiar información sin recurrir a la transcripción manual. Este tipo de aplicaciones ya tienen un uso generalizado en los servicios bancarios, de gestión de bibliotecas, médicos y otros. Un concepto de mayor alcance sería almacenar estos datos en "cajas fuertes de datos digitales" en la Internet. Una vez autorizados por el titular, los usuarios de la información (como las organizaciones censales) podrían recuperar de ellas la información necesaria.

En el mejor de los casos, el uso de la biometría o de las tarjetas inteligentes personales en los registros civiles o los censos hasta ahora ha

sido experimental y las cajas fuertes de datos digitales no pasan de ser una idea muy nueva. No obstante, es fácil prever que estas y otras técnicas similares, con sus distintas repercusiones, se convertirán cada vez más en objeto de debate en un futuro no tan lejano. Forman parte de lo que se ha dado en llamar la “informática omnipresente”, es decir, una presencia cada vez mayor del poder de la computación, y de los sensores y controles asociados, en la vida cotidiana. Las organizaciones de estadística deben participar en este debate a fin de asegurar que las nuevas tecnologías y sus normas tomen en cuenta sus necesidades.

En algunos países, los censos puerta a puerta han sido reemplazados por lo que se denomina censos “administrativos” o “virtuales” (Laan, 2001). Este sistema puede hacer necesarias una inspección y una fusión integrales de varios registros para obtener el universo nacional de viviendas y personas. También en este caso, sería ventajoso que la oficina de estadística participe en la definición y el mantenimiento de los registros más importantes. En otros países, el “formulario corto”, que incluye las preguntas que deberán hacerse a todas las personas, se ha reducido al mínimo.

En ambos casos (los censos administrativos y el formulario corto mínimo) la información adicional generalmente se obtiene mediante encuestas muestrales. Estos métodos tienen fundamentos metodológicos comunes. Las buenas prácticas estadísticas establecen el uso de métodos de muestreo siempre y cuando el universo subyacente sea suficientemente conocido. En las próximas rondas censales, las técnicas de muestreo serán cada vez más importantes y de ahí la importancia de que los estadísticos expliquen sus metodologías a todo el mundo.

X. CONCLUSIONES

Las nuevas tecnologías se incorporan a la labor censal, pero no siempre con la rapidez y la amplitud que cabría esperar. Los encargados de la planificación de los censos deben ser prudentes, pues saben que las soluciones que proponen deben ser adecuadas desde la primera vez que se aplican. No obstante, muchas veces las innovaciones se convierten en prácticas establecidas, como el trazado digital de mapas, el reconocimiento inteligente de caracteres y las publicaciones electrónicas. La Internet se ha convertido en un medio esencial de divulgación de la información y adquirirá cada vez más importancia en las actividades de recopilación de datos.

Las condiciones en que se realizan los censos varían mucho de un país a otro. Aun para la mayoría de las subareas, no puede afirmarse que

un enfoque tecnológico determinado sea el mejor. El conocimiento de la tecnología, el sentido común, los enfoques metodológicos y un tiempo más que suficiente de preparación son los requisitos fundamentales para los encargados de la planificación de los censos.

La tecnología empleada en los censos se modifica mucho más rápidamente que los métodos y principios estadísticos subyacentes. Las nuevas tecnologías nunca deben poner en peligro la continuidad de los sistemas existentes de información y, siempre que sea posible, deben reforzarlos.

La externalización plantea muchos problemas, incluidos los relativos a la confidencialidad, pero puede permitir lograr ciertas economías y resolver algunos cuellos de botella. Una vez más, cabe aclarar que la situación del país (incluida la capacidad de gestión de la oficina de censos) será un factor determinante de la viabilidad de una alternativa propuesta. La solución es más evidente para las operaciones especiales, que se realizan una sola vez (como el ingreso de los datos censales), que para las tareas permanentes, como la gestión del sitio web. Cuando la externalización pueda aportar ventajas pero existan obstáculos burocráticos, éstos deben eliminarse.

En la actual ronda de censos se observa una considerable evolución en materia tecnológica con respecto al ciclo anterior. Cabe suponer que en la próximas rondas las diferencias serán aún mayores.

BIBLIOGRAFÍA

- Basset, P. y A. Stoyka (1996), "Statistics Canada's aggregate output database – CANSIM II", *Proceedings of the Conference on Output Databases*, Voorburg.
- Blum, Olivia (1997), "Editing and coding module", *New Census Technologies: The Israeli Experience. Proceedings of the Euro-Med Workshop*, marzo.
- Dekker, Arij (1997), *Data Processing for Demographic Censuses and Surveys, with Special Emphasis on Methods Applicable to Developing Country Environments*, La Haya, Fondo de Población de las Naciones Unidas (FNUAP)/Instituto Demográfico Interdisciplinario Holandés (NIDI).
- Deming, W. Edwards (1986), *Out of the Crisis*, Boston, Massachusetts, Instituto de Tecnología de Massachusetts (MIT), Centro de Estudios Avanzados de Ingeniería.
- ____ (1994), "Computer Methods in Population Census Data Processing", *International Statistical Review*, vol. 62, N° 1.
- Dopita, Patricia (1999), *Population Census Evaluation, 1996 Census Data Quality: Occupation*, Belconnen ACT, Oficina de Estadística de Australia.

- ESCAP (Economic and Social Commission for Asia and the Pacific) (2001), "Report on the Workshop on Population Data Analysis, Storage and Dissemination Technologies" (Bangkok, 27 al 30 de marzo) (<http://www.unescap.org/stat/pop-it/pop-wdt/pop-wdt.htm>).
- Figueiredo, José y Ana Lucas (1999), *Potentials and Pitfalls of INE-P IS/IT Strategy on the Past Ten Years. Proceedings of the Strategic Reflection Colloquium on IT Issues for Statistics*, Luxemburgo, Oficina de Estadística de las Comunidades Europeas (EUROSTAT)/Instituto Nacional de Estadística de Portugal, septiembre.
- FNUAP (Fondo de Población de las Naciones Unidas) (2000), *Report of Joint Interagency Coordinating Committee on Censuses for sub-Sahara Africa and PARIS 21 Census Task Force Meetings*, Luxemburgo, Oficina de Estadística de las Comunidades Europeas (EUROSTAT), octubre.
- Giessing, Sarah (1999), "Transferable software for automated secondary cell suppression", documento presentado en el Seminario sobre Intercambio de Tecnología y Conocimientos Especializados (ETK), Praga, Oficina de Estadística de las Comunidades Europeas (EUROSTAT).
- Haug, Werner y Marco Buscher (2000), *E-census, the Swiss Census 2000 on the Internet*, Taller del Instituto Nacional de Estadística y Estudios Económicos y la Oficina de Estadística de las Comunidades Europeas (INSEE/EUROSTAT), "Los Censos después del 2001" (París, 20 al 21 de noviembre).
- Jacob, Michel y Jean-François Royer (1999), "Le recensement de la population de 1999", *Les actualités du Conseil national de l'information statistique*, N° 30, enero.
- Jain, Anil y otros (comps.) (1999), "Biometrics: personal identification in networked society", *Kluwer International Series in Engineering and Computer Science*, vol. 479, Dordrecht, Kluwer Academic Publishers.
- Kambayashi, Yahiko, Mukesh Mohania y A. Min Tjoa (comps.) (2000), "Data Warehousing and Knowledge Discovery", documento presentado en la Segunda Conferencia Internacional, DaWaK 2000 (Londres, 4 al 6 de septiembre).
- Keller, Wouter (1999), *Preparing for a New Era in Statistical Processing: How new technologies and methodologies will effect statistical processes and their organisation, Proceedings of the strategic reflection colloquium on IT issues for statistics*, Luxemburgo, Oficina de Estadística de las Comunidades Europeas (EUROSTAT), septiembre.
- Keller, Wouter y Ad Willeboordse (2000), "Statistical Processing in the Internet Era: the Dutch View", documento presentado en la "Conferencia sobre las Redes de Estadística para Mejorar el Cumplimiento de las Normas y la Calidad de las Operaciones en Europea" (Radenci, 13 al 15 de noviembre) (<http://www.sigov.si/zrs>).
- Laan, Paul van der y Peter Everaers (2001), "The Dutch Virtual Census", documento presentado en la "Sesión 53 de la Asamblea 66 del Instituto Internacional de Estadística (ISI)", Seúl, inédito.

- Meyer, Eric y Pascal Rivière (1997), *SICORE, un outil et une méthode pour le chiffrement automatique à l'INSEE*, París, International Blaise User Group.
- Naciones Unidas, (1998), "Principles and Recommendations for Population and Housing Censuses, Revision 1", Statistical Papers Series M, N° 67/Rev.1, Nueva York, Departamento de Asuntos Económicos y Sociales, División de Estadística.
- Oficina del Censo de los Estados Unidos (2001), "Census Bureau breaks new Ground with Release of DVD Products", Oficina de Información Pública, 6 de febrero.
- Prewitt, Kenneth (2000), "Prepared Statement before the Subcommittee on the Census", Cámara de Representantes de los Estados Unidos, Comité sobre reforma Gubernamental, 8 de marzo.
- Statistics Sweden/EUROSTAT (Oficina de Estadística de las Comunidades Europeas) (2001), Q2001 – Conferencia Internacional sobre la Calidad de las Estadísticas Oficiales (Estocolmo, 14 al 15 de mayo) (<http://www.q2001.scb.se>).
- Whitford, David y Jennifer Reichert (2001), "Quality Assurance Challenges in the United States Census 2000", documento presentado en Q2001 – Conferencia Internacional sobre la Calidad de las Estadísticas Oficiales (Estocolmo, 14 al 15 de mayo).
- Willenborg, L. y T. de Waal (2001), *Elements of Statistical Disclosure Control*. Springer Verlag, Hamburgo.

GLOSARIO

- Almacén de datos*: el capital de datos de las empresas o las instituciones, almacenado y administrado de manera de facilitar el acceso a ellos y su análisis.
- Caja fuerte de datos digital*: un espacio en la Internet donde los ciudadanos pueden almacenar sus datos personales en forma segura y, en algunos casos, permitir el acceso a ellos.
- CASI (entrevistas asistidas por computadora)*: una técnica mediante la cual los entrevistados llenan por su cuenta cuestionarios electrónicos, asistidos exclusivamente por programas de computación diseñados especialmente para ese fin.
- CATI (entrevistas telefónicas asistidas por computadora)*: los encuestados responden a las preguntas por teléfono mientras los encuestadores ingresan las respuestas directamente a la computadora.
- CIU*: Clasificación Industrial Internacional Uniforme de todas las actividades económicas.
- CIUO*: Clasificación Internacional Uniforme de Ocupaciones.
- Codificación asistida por computadora*: actividades de codificación en que las personas adoptan las decisiones y las computadoras prestan la asistencia necesaria.

Codificación automática: la conversión, mediante computadora, de textos verbales en los códigos correspondientes, sin intervención del ser humano.

Comunicación punto a punto: (en el sentido empleado en este texto) comunicación de datos electrónicos mediante conexiones directas, sin pasar por la Internet.

Control de la confidencialidad de las estadísticas: el conjunto de medidas que prohíben el acceso no autorizado a información estadística confidencial.

Cubo de datos: estructura multidimensional para almacenar información estadística.

DVD: disco de vídeo digital; sucesor del CD-ROM, de mayor capacidad.

Externalización: delegación de una actividad (o parte de ella) a un subcontratista.

Garantía de la calidad: un patrón calificado y sistemático de todas las acciones requeridas para otorgar la confianza necesaria de que un producto cumplirá ciertas normas establecidas.

Huella de voz: un modelo digital almacenado de la voz de una persona, empleado con fines de identificación.

Identificación biométrica: la identificación de personas a través de una o más de sus características físicas.

Informática omnipresente: la omnipresencia de la capacidad informática y de sensores y controles asociados en la vida cotidiana.

Inteligencia artificial, redes neuronales, lógica difusa: diversas formas de técnicas de software innovadoras, que a menudo se basan en métodos deterministas (heurísticos).

Lenguaje orientado a objetos: lenguajes de programación para computadoras que asignan códigos a objetos y clases. Difiere de los lenguajes de procedimientos más monolíticos.

Metainformación: información auxiliar que aclara las cifras estadísticas (definiciones, normas, unidades, métodos de recopilación, etc.).

Modelo de almacenamiento relacional: actualmente el modelo de datos más utilizado para sistemas de bases de datos multipropósito. Los fundamentos teóricos fueron formulados por E.F. Codd.

NACE: Clasificación Industrial General de Actividades Económicas de las Comunidades Europeas; clasificación estadística de las actividades económicas empleada por la Unión Europea.

RIC (reconocimiento inteligente de caracteres): el arte de interpretar caracteres escritos o impresos mediante el escaneo de imágenes y el análisis computarizado; solía denominarse reconocimiento óptico de caracteres, cuando el papel de los dispositivos de reconocimiento era menos importante.

Robot de conocimientos (knowbot): agente inteligente que reúne información de la Internet; más específico que los motores de búsqueda.

SIG (sistema de información geográfica): un sistema de información cuyo objeto es adquirir, almacenar, actualizar, manipular, analizar y exhibir todas las formas de información vinculadas con una ubicación geográfica.

Sistema de posicionamiento mundial: hoy por hoy, instrumento muy común que indica la posición de la persona o vehículo que lo transporta.

Tablero de anuncios: un servicio de información digital que a menudo opera en forma independiente de la Internet.

Tarjeta inteligente: tarjeta electrónica con un microcircuito incorporado, capaz de desempeñar funciones que van mucho más allá de una simple memoria.

Tecnología de alimentación activa: el uso de la Internet para enviar información determinada pero no solicitada a determinadas direcciones de correo electrónico.

Telefonía satelital: comunicaciones telefónicas mediante satélites geoestacionarios, sin estaciones de retransmisión terrestres.

Teleobservación: monitoreo a distancia desde aeronaves o satélites terrestres.

TIC: tecnologías de la información y las comunicaciones.

Transacciones en cajeros automáticos: una transacción realizada mediante un cajero automático o dispensador de dinero.