

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 CARIBENO 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

¿Vale la pena hacer encuestas de post-empadronamiento censal? <i>David C. Whitford y Jeremiah P. Banda</i>	259
Preservación de archivos con documentos y microdatos censales y aumento de los grupos de gestión. <i>Wendy L. Thomas y Robert McCaa</i>	303
El software Redatam para divulgación y análisis de datos censales. <i>Dirk Jaspers Faijer y Serge Poulard</i>	321

**PRESERVACIÓN DE ARCHIVOS CON DOCUMENTOS
Y MICRODATOS CENSALES Y AUMENTO
DE LOS GRUPOS DE GESTIÓN***

Wendy L. Thomas y Robert McCaa*****
University of Minnesota Population Center

RESUMEN

Cuando los datos y documentos censales han sido bien preservados permiten una efectiva recopilación, difusión, planificación y un buen uso futuro. En la era electrónica se corre el riesgo de que tanto la documentación como los microdatos no estén bien preservados, es decir, útiles para su uso; de hecho, se pueden perder o quedar ilegibles debido a la obsolescencia tecnológica o a la falta de cuidado. En este trabajo se analizan temas relacionados con la preservación a largo plazo: qué preservar, cómo determinar el valor futuro y de qué manera una buena política en tal sentido puede hacer que crezca el número de grupos de interés. El valor potencial del censo puede aumentar sustancialmente, en particular el estudio de los procesos sociales y demográficos que se produjeron a través del tiempo. Ello se obtiene mediante el uso continuo de los microdatos censales, como sucede con los proyectos IPUMS (estadounidense e internacional), SAR (Reino Unido) y otros.

* Este documento fue presentado en la Reunión del Grupo de Expertos del Simposio sobre el Examen Mundial de la Ronda 2000 de los Censos de Población y Vivienda. División de Estadística, Secretaría de las Naciones Unidas, Nueva York, 7 al 10 de agosto de 2001.

** wlt@pop.umn.edu.

*** rmccaa@umn.edu.

ABSTRACT

Proper preservation of census data and documents contributes to effective processes of collection dissemination and planning, and the future use of the censuses. In the electronic age there is the risk that both documents and microdata may not be well preserved, that is to say, may no longer be useful; they may be lost or become illegible owing to technological obsolescence or a lack of care. This paper reviews topics related to long-term preservation: what to preserve, how to determine future value and in what way a good policy in that respect can increase the number of interest groups. The potential value of the census may increase substantially, especially for the study of social and demographic processes that have taken place over time. Such studies rely on the continuous use of census microdata, as in the case of IPUMS (United States and international), SARs (United Kingdom) and other projects.

RÉSUMÉ

Des données et des documents censitaires bien préservés permettent une collecte, une diffusion et une planification efficaces, ainsi qu'une bonne utilisation dans l'avenir. L'ère électronique implique certains risques quant à la préservation de la documentation et des microdonnées en termes de leur utilisation : ces données peuvent effectivement se perdre ou devenir illisibles en raison d'une technologie obsolète ou d'un manque d'entretien. Cette étude porte sur l'analyse de différents aspects de la préservation à long terme : quels sont les éléments à préserver, comment déterminer la valeur future et comment des actions judicieuses en ce sens peuvent contribuer à l'émergence de nouveaux groupes d'intérêt. La valeur potentielle du recensement peut augmenter considérablement, notamment l'étude des processus sociaux et démographiques qui ont évolué au fil du temps, et ce grâce à l'utilisation permanente des microdonnées censitaires, comme cela est le cas dans les projets IPUMS (des Etats-Unis et international), SAR (Royaume-Uni), etc.

I. INTRODUCCIÓN

La preservación de la documentación censal debe ser considerada al comienzo del ciclo de actividades censales, pues contribuye a una efectiva recopilación, difusión, planificación y uso futuro de los censos. La capacidad para aprender de procesos pasados, identificar estrategias para un censo exitoso, conservar y recurrir a actividades y estructuras centrales de censos anteriores, y aplicar eficazmente los datos censales a problemas actuales y futuros depende de la preservación de los datos censales y la documentación relacionada con la recopilación y procesamiento de esos datos.

Es fácil determinar lo que se debe preservar y el modo en que debe ser preservado si se contara con un mundo ideal de recursos ilimitados. Desafortunadamente, éste no es el caso y, aun en los países más ricos, el costo de la preservación y de los temas que giran en ese entorno tienen un profundo impacto en la documentación y en el formato en que deben ser preservados. El objetivo de este trabajo es analizar los tipos de datos y documentación acumulados durante el proceso censal y explorar los beneficios que genera la preservación de este tipo de documentos para futuros censos o usuarios de datos, el aseguramiento de formatos de preservación, adecuación e identificación de grupos de interés que pueden constituir una fuerza eficaz que abogue por la preservación de estos documentos.

La clasificación de la documentación para preservación, en términos de su impacto futuro y de la anticipación de su uso, es útil para identificar las ventajas comparativas de las decisiones que toma cada país.

Al empalmar este tipo de listas de documentación con un inventario de la tecnología disponible, de personal y de conocimientos dentro de un país para procesar la documentación que se desea preservar, los gobiernos contarán con información que les permitirá tomar decisiones bien fundadas. El uso de un cuestionario que genere información sobre infraestructura disponible de preservación dentro de un país también puede brindar opciones de servicios cooperativos o perfilar tecnologías que se adecuan a diversas situaciones. La capacidad para determinar no sólo lo que se preservará sino también lo que no lo será —basada en una comprensión del impacto a largo plazo que tendrá la información incluida en el documento—, es crucial para desarrollar una política de preservación a largo plazo.

II. PRESERVACIÓN A LARGO PLAZO DE LOS DATOS Y LA DOCUMENTACIÓN

a) Definición de preservación a largo plazo

La preservación a largo plazo asume un nuevo significado con los registros electrónicos. “Archivar” es un término utilizado tanto por especialistas en computación/informática como por archivistas; sin embargo, para estos dos grupos el término tiene distintos significados. “Archivar” en el mundo de la computación significa almacenamiento inactivo o memoria indirecta (“off-line”). Para los archivistas, “archivar” significa preservar un registro de información en un formato que es independiente de su medio de producción y protegerlo contra pérdida, modificación o deterioro.

Para los archivistas, un registro electrónico bien preservado tiene las siguientes características. (Dollar, 2000, 47-57):

- **Legible.** Es decir, no está dañado y la secuencia de bits puede ser procesada ya sea por la máquina que la creó, la que la está almacenando, o aquella en la que será almacenada.
- **Inteligible.** Que tiene suficientes metadatos para interpretar los unos y ceros de la imagen en el mapa binario. En otras palabras, información sobre el algoritmo de compresión y el orden de los bytes. Es similar a la extensión de archivo TXT que denota un archivo de texto ASCII de 7 bits. Si carece de este nivel básico de metadatos, el registro es ininteligible para todos los efectos prácticos.
- **Identificable.** Es decir, puede ser localizado mediante una identificación o atributo único.
- **Encapsulado** de modo tal que toda la información del registro (sus metadatos y vínculos) exista como una sola entidad lógica o física.
- **Comprensible** a través del suministro de metadatos completos.
- **Regenerable** en cuanto al contenido lógico, físico e intelectual.
- **Registros auténticos.** “La ciencia de archivar define a los registros auténticos como lo que aparentan ser: registros confiables que con el correr del tiempo no se han modificado, cambiado o corrompido” (Dollar, 2000, 54).

Es importante que este concepto de preservación sea tenido en cuenta al momento de determinar el valor de preservar ciertos registros censales y los costos de distribución, almacenamiento y preservación a largo plazo.

b) El valor de la preservación

Mucho se ha escrito sobre la importancia de organizar y coordinar la realización de censos dentro y entre países (Naciones Unidas, 2000). Este proceso cuenta con el apoyo y asistencia de numerosas agencias intergubernamentales y no gubernamentales. Se ha puesto énfasis en la planificación, recopilación de datos, metodologías, preparación del producto y difusión. El valor de un buen programa censal reside no sólo en preservar datos reales, metadatos y resultados para su uso futuro sino también en servir de apoyo para futuros censos y actividades estadísticas.

Dada la naturaleza periódica de los censos, la conservación de registros sobre la manera en que se llevaron a cabo ciertas actividades específicas puede ser de ayuda en futuros procesos censales de un país y permitir que las agencias adquieran experiencia basada en procesos y estrategias pasadas. Esto reviste particular importancia para aquellos países que no tienen una oficina permanente para el censo. Los registros cuidadosamente seleccionados y preservados brindan información detallada sobre el proceso de planificación y sobre las especificaciones referidas al proceso de recopilación, y explican por qué se tomaron algunas decisiones y cuán efectivas resultaron ciertas actividades. De hecho, estos tipos de procesos y enfoques específicos nacionales son los que podrían ayudar a incorporar y aprovechar actividades y estructuras exitosas.

La preservación y difusión de información sobre calidad de datos y evaluación del proceso es valiosa para futuras actividades censales y crucial para que el uso de datos censales esté bien fundado. Si se entrega información sobre la confiabilidad, limitaciones y validez de los datos finales, los usuarios comprenderán que cualquier cambio de procedimiento puede afectar cualquier análisis que se desee realizar. Éste es el tipo de información que debería ser encapsulada en el proceso de preservación a través de enlaces lógicos o físicos entre los datos censales y los metadatos de procedimiento.

c) Costos de la preservación

El costo de la preservación es un tema importante en todos los países. Los recientes debates sobre conservación del censo 2000 de los Estados Unidos de Norteamérica generaron numerosas respuestas de varios grupos de interés referidas a la preservación de los formularios originales y de los resultados de procesos intermedios. El costo de preservación de formularios con enumeración original en varios formatos y el costo asociado en que se

incurre para hacerlos identificables por futuros usuarios fueron factores clave en la negociación de un plazo final de conservación.

En aquellos países que no poseen oficinas permanentes de censos y/o instalaciones permanentes para archivos nacionales, el costo de preservación es un gran problema. Si estos costos se tienen en cuenta con anticipación y se los incluye en el análisis de los costos generales del censo, se podrán encontrar opciones adicionales para la asignación de fondos. Por ejemplo, la forma en que los datos censales se obtienen y preparan para su difusión puede reducir el costo que implica crear un registro de preservación de buena calidad. Además, la captura y conservación de información sobre procedimientos a medida que la misma se genera y la creación de enlaces lógicos o físicos con recopilaciones de datos emergentes aumentan la posibilidad de la preservación y, al mismo tiempo, reducen el costo de tener que reconstruir valiosa información de metadatos.

Un análisis previo de los costos y del valor futuro de preservar información permite tomar decisiones bien fundadas y da la oportunidad de discutir en forma oportuna posibilidades de preservación a largo plazo.

III. DETERMINACIÓN DE LO QUE SE DEBE PRESERVAR

a) Preservación de productos

Los elementos esenciales de cualquier censo, en términos de preservación, son los datos y la documentación básica. La forma en que esos datos se identifican y definen varía de país en país. Los temas de confidencialidad y seguridad desempeñan un papel fundamental para determinar no sólo quién debería tener acceso a los microdatos y a los formularios de enumeración sino también si esa información debería ser conservada. Una mayor disponibilidad de microdatos aumenta la posibilidad de que estos datos sean preservados.

Son cada vez más los países que proporcionan acceso a microdatos a través de: muestras públicas, muestras científicas (restringidas a unos pocos proyectos cuidadosamente seleccionados) y enclaves de datos donde el usuario trabaja en un sitio seguro y el producto es estrechamente controlado. Desde 1985 hasta 1994, de 153 países con un millón o más habitantes, 134 realizaron censos en la ronda de 1990; se contó al 94% de la población mundial. Cincuenta y cuatro países permitieron que los investigadores tuviesen acceso a muestras anónimas de censos de población y vivienda. Otros restringieron el acceso a un solo investigador o centro de

investigación, pero lo que es notable en la ronda de 1990 es no sólo la globalización del censo sino la creciente aceptación de muestras anónimas como instrumentos estadísticos, tendencias que siguen vigentes en la ronda de censos del 2000 (1995-2004).

Por ejemplo, el enfoque utilizado en las Naciones Unidas de suministrar muestras públicas cuyos tamaños variaban entre 1% y 15% en varios tipos de zonas sirve de soporte a una amplia gama de investigaciones tanto a nivel local como nacional. Además, la divulgación de datos (después de haber estado restringidos durante 73 años) ha generado una cantidad de proyectos destinados a que el público tenga acceso a los mismos en formato digital. El más notable es el proyecto Integrated Public Use Microsample (IPUMS). Este proyecto, que se inició en 1992 en la Universidad de Minnesota, incluye 65 millones de registros de microdatos de los Estados Unidos de América. Concebido por Steven Ruggles —director fundador del Minnesota Population Center— y financiado por la National Science Foundation/National Institute of Health, IPUMS es parte de los censos decenales de los Estados Unidos de América, que datan desde 1850 a 1990. La primera versión de la base de datos IPUMS se difundió en cinta magnética en 1993 y por Internet en 1995. Gracias a la expansión de Internet, el problema de distribución de datos se resolvió fácilmente mediante un motor de difusión de datos a través de un sitio web (<http://www.ipums.org>). La base de datos IPUMS, distribuida en forma gratuita por Internet, se convirtió rápidamente en una de las tres fuentes de datos más frecuentemente citadas en los estudios de investigación demográfica de los Estados Unidos de América.

Con fondos suministrados mayormente por la National Science Foundation, en octubre de 1999 se inauguró un proyecto mundial, denominado IPUMS-International. El consorcio IPUMS-International, que cuenta con la colaboración de equipos nacionales de investigadores, propone integrar microdatos censales en más de una docena de países, por lo menos uno de cada continente. Se incluirán en la base de datos microdatos históricos de censos de Canadá, Noruega, Gran Bretaña, Argentina y Costa Rica y también de los Estados Unidos de América. Los microdatos actuales de Colombia y los Estados Unidos de América se integrarán con los de Francia, Brasil, Méjico, Vietnam, Kenya, Gran Bretaña, Hungría, España y otros. Basándose en un prototipo desarrollado con la cooperación del Departamento Administrativo Nacional de Estadística de Colombia (DANE), se están formando equipos nacionales de usuarios experimentados en el uso de datos censales para que asesoren sobre la manera de unificar los conceptos nacionales de censos utilizando normas internacionales.

Varios países están creando muestras de uso público para incrementar el acceso a los microdatos. Programas del tipo Integrated Microcomputer Processing System (IMPS) y su sucesor CSPro, un sistema de procesamiento de datos censales y de encuestas desarrollado por el International Statistical Programs Center del U.S. Bureau of the Census facilita la difusión de muestras de microdatos suministrando herramientas para realizar tabulaciones cruzadas, producir mapotecas digitales y otros análisis básicos con lo cual se reduce el costo de producción de estos productos en cada país.

Entre los países que distribuyen muestras de microdatos de uso público se encuentra Vietnam, que dio a conocer una muestra (3%) del Censo de Población y Vivienda de 1990 con la intención de producir más adelante una muestra completa del 100%. México difundió un 10% de una muestra diseñada para generar información valiosa a nivel de municipalidades de 100 000 o más habitantes. Francia difundió 5% de las muestras de 1961-1990. Del mismo modo, la Central Bureau Statistics de Kenya ha preparado una mega muestra de la enumeración 1999 (con una densidad máxima del 20%) a fin de completar su impresionante serie de muestras de 1969, 1979 y 1989.

Estas recopilaciones no sólo suministran datos en un formato que se puede preservar sino que incluyen una gama de metadatos. La documentación es sumamente completa e incluye detalles sobre cada aspecto del censo, desde los preparativos iniciales hasta la publicación final de los cuadros. El debate del muestreo es particularmente notable.

Un creciente número de países ofrece datos en formato REDATAM (desarrollado por la División de Población de la CEPAL-Centro Latinoamericano y Caribeño de Demografía (CELADE)) como una forma de almacenar microdatos que pueden utilizar aquellos investigadores y administradores que necesitan estadísticas para áreas pequeñas.

REDATAM “**RE**cuperación de **DAT**os para **Á**reas pequeñas por **M**icrocomputadores” fue originalmente concebido como un programa de computación para la recuperación de datos a bajo costo y se ha convertido en un concepto que comprende un formato patentado de base de datos y un sistema para el desarrollo de software. El formato busca asegurar datos sensibles manteniendo al mismo tiempo la invalorable flexibilidad del acceso a los microdatos. También dispone de un servicio web que beneficia a organizaciones nacionales que se resisten a proporcionar datos pero que están dispuestas a brindar al público acceso a los datos y/o acceso privilegiado a usuarios seleccionados. El programa está disponible en forma gratuita en Internet. REDATAM ha sido desarrollado durante las dos últimas

décadas gracias al apoyo financiero de diversas organizaciones internacionales (CEPAL-Naciones Unidas, FNUAD, Gobierno de Canadá a través de CIDA e IDRC, IDB y otras) (<http://www.cepal.cl/celade>).

Países con ronda de censos de 1990 en REDATAM:

América Latina: Argentina, Brasil, Chile, Colombia, República Dominicana, Guatemala, Honduras, Nicaragua, Paraguay, Suriname, Uruguay, Venezuela, y el Caribe de habla inglesa.

Asia: Camboya* y Corea del Norte*

África: Benin, Burkina Faso, Burundi, Camerún, Egipto, Gabón, Ghana, Kenya, Madagascar, Malí, Nigeria, Ruanda*, Seychelles, y Zimbabwe*.

* Base de datos con 100% de microdatos de población.

Si bien estos archivos de microdatos no están en formato de archivo en sentido estricto, han sido capturados de forma tal que el organismo que los generó puede producir un archivo en formato ASCII con metadatos estructurales completos físicamente encapsulados para asegurar su futura comprensión. Es importante que formatos como REDATAM no sean considerados como formatos de archivo a largo plazo. El problema que se presenta cuando no se crea una copia de archivo y los registros se guardan en un formato patentado es el costo de llevar esa información a otro formato. Los formatos patentados pronto se convierten en formatos heredados, cuya edad, dependencia de sistemas, idiomas o hardware los hace difíciles, costosos y a veces imposibles de traspasar.

b) Preservación del proceso

Diversos manuales sobre realización y administración de censos nacionales proporcionan listas detalladas de procedimientos y procesos. Este tipo de información y los detalles sobre ciertos enfoques y metodologías son necesarios para interpretar en forma precisa los datos resultantes. También sería conveniente considerar los tipos de información de proceso, que a veces se pasan por alto, para preservar la memoria institucional. Esto implica registrar y preservar las razones y formas del proceso censal. Es más eficiente en función de los costos utilizar esta información mientras se toman las decisiones que recuperarla más adelante. Se debe prestar especial atención al hecho que dicha información debería ser utilizada en formato no patentado para evitar que la misma se pierda debido a los costos de traspaso.

Todo ciclo censal consiste de cuatro fases (Naciones Unidas, 2000).

- Preparación
- Actividades sobre el terreno
- Procesamiento de datos
- Evaluación

Entonces, para cada fase reviste particular interés la siguiente documentación:

- Informes sobre procedimientos y métodos.
- Comparación de conceptos y procedimientos con censos anteriores y normas internacionales vigentes.
- Informes de evaluación para cada ciclo censal y documentos más importantes en los cuales se basan los informes.
- Libros de registros (desde el del jefe del censo hasta el registro del enumerador), aunque éstos pueden ser demasiado primarios para ser difundidos ampliamente.

La etapa final consiste en documentar los datos difundidos y la respectiva documentación y notas.

IV. CÓMO DETERMINAR EL VALOR FUTURO

a) Quiénes forman los grupos de interés

Según lo ya expresado aún continúan surgiendo normas para permitir el acceso a microdatos censales; el principal problema relacionado con su preservación y acceso no es técnico sino político. A medida que el acceso y uso de los datos censales aumenta, la naturaleza y complejidad de los grupos de interés también aumentan; éstos se encuentran en el sector gubernamental, no gubernamental, académico, comercial y de usuarios nuevos. Si bien siempre habrá intereses que compiten por lo que debería ser conservado, los compartidos por diversos grupos los que ayudarán a identificar la documentación que será preservada a largo plazo. Si se celebran consultas con estos grupos antes de iniciar el proceso, aumentará la posibilidad de obtener y conservar financiamiento para la preservación a largo plazo.

b) Impacto futuro y uso anticipado

En términos de patrones de impacto futuro y uso de datos censales, el mayor potencial radica en preservar y mantener el acceso a microdatos.

Los datos censales se usan para analizar problemas específicos de orden social, económico y demográfico, cuya naturaleza cambia constantemente. La capacidad para crear agregados comparables a lo largo del tiempo o para identificar nuevas áreas geográficas o definiciones emergentes depende solamente de la conservación de archivos de microdatos. Éste es el caso de las estadísticas para áreas pequeñas que no se pueden obtener de agregados de áreas más extensas.

Las tabulaciones específicas tienen claras dificultades. El ciclo censal y la producción de agregados estadísticos significan que las preguntas formuladas y los cuadros creados a menudo reflejan inquietudes e intereses que datan de cinco a diez años previos a la publicación. Es probable que las nuevas tabulaciones no sean comparables con las tabulaciones de censos previos debido a cambios en la cohorte o en los agrupamientos de clasificación, en el universo del cuadro o en la definición. Sin acceso a microdatos, sin importar cuán asegurado esté, los investigadores y analistas tienen pocas opciones. Además, los microdatos se prestan para la investigación y el discurso académico y aumentan el valor derivado de un censo individual.

Consideremos el ejemplo de Canadá: en la década de 1970, el National Statistical Services comenzó a difundir grandes cantidades de muestras de microdatos censales. En Canadá, la modificación de la Ley de Estadísticas llevada a cabo en 1971 permitió la difusión pública de microdatos no confidenciales (Tambay y White, 2001). A partir de la década de 1970, Statistics Canada, con su serie de enumeraciones quinquenales, emitió en forma regular muestras de microdatos censales. Hasta el año 1996 los investigadores tenían que solicitar las muestras en forma individual y la distribución estaba muy restringida. En ese año se puso en práctica una propuesta de difusión de datos para que las universidades canadienses pudiesen suministrar muestras de microdatos a los investigadores y alumnos. Como resultado de ello se produjo un desarrollo explosivo de trabajos de investigación. Si bien antes de que se permitiera la difusión sólo cinco o diez especialistas al año podían obtener muestras de microdatos, después de haberse autorizado la difusión se podía tener acceso a una sola muestra cientos de veces al mes en una universidad muy importante. Dada la profusión de proveedores, actualmente es imposible compilar estadísticas de uso, mientras que antes el organismo registraba el nombre de cada usuario. Como el uso de microdatos censales se ha extendido dentro de las aulas universitarias, los especialistas canadienses están enseñando a una generación de ciudadanos más jóvenes la utilidad del censo y están democratizando el

acceso a los datos censales (Lisa Dillon, comunicación privada, 21 de abril de 2001).

En el Reino Unido se construyeron muestras de censos de uso público llamadas SAR (Samples of Anonymized Records) para la enumeración de 1991 con una densidad de muestras del 2.0% de registros individuales y del 0.5% de viviendas. No se identificaron unidades administrativas con menos de 120 000 habitantes. A pesar de la pequeña densidad de las muestras y de la falta de detalles geográficos, se produjo un desarrollo explosivo de trabajos de investigación que utilizaban las SAR. Durante los seis años posteriores a la primer difusión de datos, se publicaron cientos de estudios. Antes de la enumeración de 2001, se realizó una nueva evaluación de los riesgos de la divulgación y en ese momento se tuvo en cuenta la varianza residual y de codificación así como las diferencias en los programas cronológicos y de codificación entre grupos de datos. Con el permiso de la Office of National Statistics se otorgó acceso privilegiado para comparar los registros de encuestas con las SAR (Dale y Elliot, en prensa). El objetivo era determinar los riesgos prácticos, en oposición a los teóricos, de identificar individuos combinando dos fuentes. Los autores consideraron que las evaluaciones anteriores sobre la posibilidad de identificar individuos exageraban los riesgos porque no tuvieron en cuenta la varianza residual, las diferencias de oportunidad y las incompatibilidades de los programas de codificación. A partir de este riguroso ejercicio, Dale y Elliot llegaron a la siguiente conclusión:

Si un usuario de una base de datos externa no tiene oportunidad de llevar a cabo una verificación, sería inútil que intente realizar este tipo de combinación. En primer lugar, es dable esperar un pequeño grado de superposición lo cual sería un gran factor de disuasión para un intruso.

Sin embargo, si se intenta combinar ambos archivos, la gran cantidad de combinaciones aparentes sería muy confusa ya que el intruso no tendría forma de verificar la identificación correcta.

c) Información sobre censos futuros

La disponibilidad de información de un proceso que es propio de un país y/o estructura organizativa puede servir de insumo en la preparación de censos futuros al brindar un panorama claro de lo que sucedió, cómo sucedió, por qué se llevó a cabo de esa manera y de los logros y dificultades. Este tipo de información reviste particular importancia para aquellos países que no tienen una oficina permanente o que cuentan con mínimo personal

permanente y que deben crear un nuevo sistema para cada censo. El suministro de documentación sobre la razón para qué los procesos y procedimientos se siguieron de cierta forma también puede ser de ayuda para los consultores técnicos internacionales, porque les brinda un panorama claro y completo de las actividades realizadas durante los censos anteriores.

V. INVENTARIO DE TECNOLOGÍA/PERSONAL/ CONOCIMIENTOS DISPONIBLES

Antes de la ronda de censos de 1990, la División de Estadística de las Naciones Unidas distribuyó un cuestionario referido a la cobertura general del censo, estructura organizativa, trabajo cartográfico, listado de casas y/o viviendas, ensayos (ensayos previos, ensayos piloto, etc.), cuestionarios del censo, enumeradores y supervisores, enumeración, muestreo, procesamiento de datos, evaluación y análisis, difusión de datos, costos y actividades futuras. Además de la información ya solicitada, las siguientes áreas de información relacionadas con la preservación a largo plazo servirían de ayuda para que los países incluyan el tema de la preservación al inicio del proceso. El reconocimiento anticipado de las necesidades y posibilidades de preservación permitirá tomar decisiones de preservación bien fundadas.

- Existencia de un archivo nacional capaz de preservar documentación digital.
- Existencia de un plan de preservación que incluya un plazo de conservación de registros.
- Tipo de documentación (resultados de datos, metadatos, documentos del proceso) que se está preservando y que se espera preservar.
- Disponibilidad de personal capacitado en preservación.
- Uso de depósitos/archivos para datos y documentación.

VI. CONCLUSIÓN

Si los microdatos censales llegan a usarse en forma extensa, es necesario resolver el tema de la confidencialidad estadística de forma tal que satisfaga a los organismos nacionales de estadística, al público y a investigadores.

En la última década Eurostat patrocinó cinco conferencias internacionales sobre el tema. Gracias en parte a estos esfuerzos, y a otros, la regla general ahora consiste en preparar muestras de microdatos para una diversidad de usuarios. De los 52 estados miembro del International Monetary Fund's General Data Dissemination System (Sistema de Difusión de Datos Generales del Fondo Monetario Internacional), unos tres de cuatro difunden muestras de microdatos censales, de un modo u otro. El establecimiento de normas internacionales de microdatos aumentará aún más la disponibilidad de muestras censales con lo que se facilitará la investigación comparativa, tanto en tiempo como en espacio. Siempre que se han adoptado políticas de difusión pública se ha producido un desarrollo explosivo en el campo de la investigación, sin que haya habido una sola instancia en que se violó o siquiera se haya supuesto que se hubiese violado la confidencialidad estadística.

Es importante comprender e incorporar este concepto de preservación ya que de ese modo se asegura que los datos censales estarán protegidos contra pérdida, modificación o deterioro.

“En este sentido, la obligación de los archivistas consiste en explicar a los especialistas en computación, a los especialistas en informática, y a los que no están familiarizados con los archivos la importancia que tiene un espacio físico o lógico, ‘independiente de su medio de producción’, donde los registros están protegidos contra pérdida, modificación y deterioro para que puedan ser utilizados como evidencia confiable todo el tiempo que sea necesario. De esto se trata archivar” (Dollar, 2000, 26).

BIBLIOGRAFÍA

- Dale, Angela y Mark Elliot (s/f), "Proposals for 2001 SARs: An assessment of disclosure risk", *Journal of the Royal Statistical Society, Series A*, Nueva York, en prensa.
- Dollar, Charles M. (2000), *Authentic Electronic Records: Strategies for Long-Term Access*, Chicago, Illinois, Cohasset Associates, Inc.
- General Statistics Office (2000) *Data and Results from the 3% Sample of The Population and Housing Census*, Hanoi, Centro de Procesamiento de Información Central, agosto.
- INEGI (Instituto Nacional de Estadística, Geografía e Informática) (2001), *Contar 2000. Sistema para la consulta de tabulados y base de datos de la muestra: XII Censo General de Población y Vivienda 2000*, Aguascalientes, México.
- Joint ECE/Eurostat Secretariat (2001), "Report of the March 2001 Work Session on Statistical Data Confidentiality. Work Session on Statistical Data Confidentiality", Skopje, Comisión Económica para Europa (CEPE)/Oficina de Estadística de las Comunidades Europeas (Eurostat), marzo.
- Naciones Unidas (2000), *Handbook on Census Management for Population and Housing Censuses, Studies in Methods*, Series F, N° 83, Nueva York, División de Estadística, Departamento de Asuntos Económicos y Sociales (DESA).
- (1992a), *Handbook Population and Housing Censuses: Part 1 Planning, Organization and Administration of Population and Housing Censuses, Studies in Methods*, Series F, N° 54, Nueva York, División de Estadística, Departamento de Asuntos Económicos y Sociales (DESA).
- (1992b), *Handbook Population and Housing Censuses: Part 2 Demographic and Social Characteristics, Studies in Methods*, Series F, N° 54, Nueva York, División de Estadística, Departamento de Asuntos Económicos y Sociales (DESA).
- (1990), *Manual on Population Census Data Processing using Microcomputers, Studies in Methods*, Series F, N° 53, Nueva York, División de Estadística, Departamento de Asuntos Económicos y Sociales (DESA).
- (1991), *Emerging Trends and Issues in Population and Housing Censuses, Studies in Methods*, Series F, N° 52, Nueva York, División de Estadística, Departamento de Asuntos Económicos y Sociales (DESA).
- Ruggles, Steven, J. David Hacker y Matthew Sobek (1995), "Order out of chaos: General design of the Integrated Public Use Microdata Series", *Historical Methods*, vol. 28.
- Ruggles, Steven y otros (2000), "IPUMS-USA: Integrated Public Use Microdata Series for the United States", *Handbook of International Historical*

Microdata for Population Research, Patricia Kelly-Hall, Robert McCaa y Gunnar Thorvaldsen (comps.), Minneapolis, Minnesota.

Tambay, Jean-Louis y Pamela White (2001), "Providing greater accessibility to survey data for analysis. Work Session on Statistical Data Confidentiality", Skopje, Joint ECE/Eurostat, marzo.