



22 JUN 2004

**INFORME SOBRE EL PROCESO DE CODIFICACION OCUPACIONAL
CENSO-2001**

C/2001 X4

Σ.2

INDICE

C. E. G.
Centro Estadístico de Servicios
DIRECCION DE DIRECCION
I. N. D. E. G.

22 JUN 2004

Introducción	pág. 3
1. Fases previas al proceso de codificación	pág. 4
El sistema de codificación informatizada	pág. 4
2. El proceso de codificación	pág. 7
El proceso de codificación de ocupaciones en el 2001	pág. 8
La codificación automática	pág. 9
La codificación semi-manual	pág. 10
La codificación en red	pág. 11
Indicadores del funcionamiento de la codificación ocupacional	pág. 12
3. Control de calidad de la información ocupacional codificada	pág. 15
El control de calidad de la codificación automática	pág. 15
El control de calidad de la codificación semi-manual	pág. 16
Estimación general del error en la codificación de ocupaciones	pág. 16
4. Codificabilidad y tratamiento de la información	pág. 19
Codificabilidad de la información ocupacional censal	pág. 19
Algunas dificultades que presentaba la información a codificar	pág. 21
5. Conclusiones	pág. 23
6. Equipo de codificadores	pág. 25

Introducción

Este informe¹ presenta las características principales del proceso de codificación de la información ocupacional relevada a partir del Censo Nacional de Población, Vivienda y Hogares 2001, así como sus resultados generales.

De esta manera se continúa la línea inaugurada en 1991, con motivo de la codificación del anterior Censo de Población y Vivienda, en cuanto a dar a conocer las metas planteadas para cada etapa, cómo se construyeron los instrumentos y procedimientos metodológicos necesarios para lograrlas y cuáles fueron los resultados de su aplicación.

Desde la conformación del Sistema Nacional de Nomenclaturas (SiNN) a mediados de los '90, ámbito donde se agrupan los equipos técnicos que desarrollan los clasificadores de rama de actividad económica, ocupación, carreras universitarias y localidad geográfica, se hizo posible la elaboración de una estrategia de trabajo común para el tratamiento de esta información, respetando las especificidades que cada temática plantea. El logro más importante de este esfuerzo común fue la creación de un Sistema de Codificación Informatizada (SiCI) capaz de afrontar el procesamiento de cada una de aquellas variables, generando las condiciones para contar con la totalidad de los datos terminados en un lapso menor que en censos anteriores. Esas condiciones hicieron viables, en lo que a la codificación ocupacional respecta, la aplicación de un nuevo clasificador a cinco dígitos² que desagrega más los grupos de ocupación que en el '91 y mide más específicamente la jerarquía y el uso de la tecnología.

En las páginas que siguen, además de la referencia al SiCI y a la elaboración del nuevo clasificador de ocupaciones, se detallan las distintas tareas involucradas en el proceso mismo de codificación. Varias fueron las innovaciones en esta instancia con respecto a 1991: el trabajo de codificación semi-manual, estructurado alrededor de una red de computadoras, lo que supuso una organización laboral distinta de la anterior; el control de calidad, desdoblado su funcionamiento en tanto hubo codificación automática y semi-manual, para señalar las dos más relevantes.

Hacia el final se evalúan los resultados de codificabilidad de la información y se los compara con los obtenidos en la codificación del Censo de 1991, evaluación que es común para el caso de la Encuesta Permanente de Hogares, pero que por primera vez puede hacerse con los resultados de un censo.

Finalmente, el ProMAEO agradece a los distintos equipos con los que se trabajó en común, así como al personal responsable de la sede Tejedor del INDEC y al equipo de informática de esa sede, por el buen clima de trabajo y la disposición para resolver los problemas técnicos que se presentaron.

¹ Este informe ha sido realizado por los integrantes del ProMAEO: Emilio Crenzel, Marcelo Esses, Patricia Höxter, Claudio La Rocca, Nora Morales, Marcelo Rosenzweig, Roberto Ruiz y Mariana Urso.

² Para conocer los avances en el nuevo instrumento clasificatorio ver Clasificador Nacional de Ocupaciones (CNO-2001), en preparación.

1. Fases previas al proceso de codificación

Teniendo en cuenta la experiencia acumulada y las conclusiones a las que se arribó en el operativo censal de 1991, varios eran los aspectos a mejorar en el nuevo censo. Tres se destacaban entre el resto: la formulación de las preguntas ocupacionales en la cédula, la clasificación de la información recolectada y el tiempo que insume procesar lo relevado para su publicación. Los objetivos buscados eran mejorar la captación de la información ocupacional, construir un instrumento clasificatorio más preciso y disminuir los tiempos de codificación.

Respecto de la captación de la información, la modificación más importante fue que a la pregunta por las *tareas desarrolladas* por el censado en su lugar de trabajo (tal como se formuló en el Censo de 1991), se agregó otra sobre el *nombre de la ocupación*, de manera tal que en el 2001 la formulación quedó del siguiente modo:

- a) "¿Cuál es el nombre de la ocupación?" (pregunta 29)
- b) "¿Qué tareas hace en ese trabajo?" (pregunta 30)

Este cambio permitió ganar en precisión y en comparabilidad con relevamientos del Instituto como la EPH y la ENGHO, entre otros. A su vez, el tratamiento conjunto de las dos preguntas mencionadas mejoró la identificación del valor de cada una de las cuatro dimensiones que conforman el instrumento clasificatorio.

En cuanto a este último, el trabajo estuvo centrado en el diseño de una nueva versión del clasificador, que superase las limitaciones operativas que presentaba la versión anterior. Fundamentalmente se amplió de 3 a 5 el número de dígitos, lo que permitió desplegar exhaustivamente las posibilidades clasificatorias, ya que cada dimensión ocuparía una posición independiente de las restantes, y mejorar la identificación de algunos grupos ocupacionales específicos³.

En relación al tercer aspecto, vinculado con el tiempo insumido por el proceso de codificación de la información censal, el INDEC tomó la decisión de desarrollar, para las preguntas abiertas, el SiCI⁴ que por su alcance y propósitos merece ser descrito en sus características principales.

El Sistema de Codificación Informatizada⁵

El desarrollo del SiCI fue impulsado y coordinado por el SiNN. Fue un trabajo que reunió a diversos sectores del INDEC a lo largo de aproximadamente 8 años. Como parte de la estructura funcional del SiNN se armó el Grupo de Aplicación formado no sólo por técnicos de la Coordinación del SiNN y del ProMAEO, sino también de la Dirección de Informática y de Metodología Estadística. Aprobada la posibilidad de diseñar el SiCI para el Censo de Población, se incorporaron a este grupo multidisciplinario técnicos de la EPH y de Desarrollo Conceptual del Censo de Población. El resultado de estos años de trabajo fue un sistema dentro del cual cada una de las variables a ser codificada tuvo su propio tratamiento.

³ Para más detalle acerca del nuevo clasificador de ocupaciones del INDEC ver Clasificador Nacional de Ocupaciones, versión 2001, en preparación.

⁴ Además de las preguntas de ocupación el SiCI se aplicaría a la rama de actividad, carrera universitaria y variables de localización geográfica.

⁵ Para mayores detalles del SiCI consultar Riestra, Mara; Lanne, Mariano "Sistema de codificación informatizada (SiCI) para operativos económicos y sociodemográficos", INDEC, Buenos Aires, febrero 2001.

Durante los primeros ejercicios de desarrollo del SiCI se pudo determinar que mientras en la rama de actividad muchos códigos se podían colocar en forma independiente de otras variables, en ocupación era muy importante establecer reglas o pautas que vincularan esta pregunta a otras, como por ejemplo, tamaño del establecimiento y categoría ocupacional. Además el método específico de codificación también sería distinto entre las dos variables: mientras que en rama de actividad se tomaría la frase completa como frase única y microprocesos por *palabra clave* dentro de una frase, en ocupación sólo se utilizaría este último procedimiento.

Las palabras clave son palabras representativas de ocupaciones de frecuencia elevada, a partir de las cuales se las puede reconocer, independientemente de la posición de las palabras en la frase. Por ejemplo, una palabra clave es "albañil". Es posible reconocer todos los casos en que aparece, ya sea que la frase completa sea "albañil", "trabajo de albañil", "albañil levanto paredes" o cualquier otra con esta palabra. Para una variable con tanta dispersión de respuestas como la ocupacional, la palabra clave sirve como factor unificador.

El inconveniente de este método es que pueden aparecer frases que incluyan la palabra clave pero refieran a otra ocupación. Siguiendo con el ejemplo anterior, las frases "peón de albañil" y "ayudante de albañil" refieren a una ocupación distinta de la de albañil (en lo que a la calificación ocupacional se refiere). Por lo tanto, hay que incluir *restricciones* y bifurcaciones dentro del proceso iniciado con la palabra clave, lo cual llevó a la conformación de una estructura arbolada con distintas resoluciones cada vez que el sistema detectara en una frase la palabra "albañil"⁶.

El sistema de codificación desarrollado para las ocupaciones supuso la localización de cerca de 100 palabras clave⁷ y el desarrollo de sus respectivas estructuras arboladas, llamadas *microprocesos*. La detección de esas palabras clave (y de sus sinónimos) y de las distintas posibilidades de resolución para cada una de ellas fue un trabajo conceptual-empírico realizado específicamente por el ProMAEO. También lo fue la secuencia en que se aplicarían los microprocesos, de acuerdo al funcionamiento del SiCI⁸.

Durante la codificación censal se incorporaron algunos nuevos microprocesos, surgidos del análisis de la información que quedaba sin codificar automáticamente. Asimismo, dada la flexibilidad del sistema, pudieron ser corregidos y completados los ya existentes.

Al diseñar el módulo de codificación ocupacional del SiCI las previsiones de asignación de código incluían tres posibles situaciones:

- **Asignación por el sistema de un código ocupacional:** se hablaba aquí de casos codificados automáticamente. Cuando el sistema asignaba un código ocupacional, éste sería el código definitivo del caso (aunque siempre se realizaron verificaciones por otros procedimientos o controles de calidad).
- **No asignación por el sistema de ningún código:** se hablaba aquí de casos no codificados. Esto podía deberse a dos motivos: o bien el caso en cuestión no

⁶ Distintas resoluciones significa en principio distintos códigos ocupacionales. Otra posibilidad es no codificar el caso, por ejemplo si en la frase hay elementos que cambien el sentido sugerido por la palabra clave tomada aisladamente.

⁷ Además, cada palabra clave iba unida a una serie de *sinónimos*, no sólo en sentido gramatical sino sobre todo ocupacional. Por ejemplo, la palabra clave "abogado" tenía como sinónimo a la palabra "escribano", ya que según el clasificador utilizado (CNO-2001) ambos casos llevan el mismo código.

⁸ El SiCI funciona secuencialmente, es decir que aplica los microprocesos en orden de prioridad. En el caso de las ocupaciones, ese orden es el mismo que se elaboró en las reglas operativas de codificación.

incluía ninguna de las casi 100 palabras clave y sus sinónimos, o bien incluía una palabra clave pero la frase contenía elementos previstos como excepción y por lo tanto no debía ser codificada automáticamente. Todos los casos no codificados debían ser resueltos por los codificadores en forma *semi-manual*⁹.

- **Asignación por el sistema de un código genérico:** estos eran códigos que si bien no estaban incluidos como tales en el Clasificador, permitían agrupar ocupaciones con algún grado de parentesco entre sí. Se hablaba aquí de casos *embolsados*, que serían resueltos posteriormente por grupos especializados de codificadores. Esta situación fue descartada durante el proceso de codificación ocupacional por razones que más adelante se referirán.

descartado

El módulo de codificación ocupacional fue integrado al SiCl el cual, además de los módulos de codificación de las otras preguntas abiertas, incluye diversas etapas de procesamiento para que la información llegue en las mejores condiciones a la codificación propiamente dicha, como por ejemplo la corrección ortográfica y la normalización y estandarización de palabras.

Las características de la aplicación del sistema en el proceso mismo de codificación y sus resultados serán presentados en el capítulo siguiente.

⁹ A diferencia de la codificación manual, hecha directamente sobre las cédulas censales, aquí la codificación sería semi-manual pues se realizaría en una pantalla de computadora, con condiciones y ventajas que serán enumeradas más adelante.

2. El proceso de codificación

Antes de exponer los aspectos operativos y los resultados del proceso de codificación conviene apuntar, brevemente, algunos datos comparativos del tratamiento de esta misma instancia de trabajo en el Censo Nacional de Población y Vivienda de 1991. En aquella oportunidad las preguntas abiertas del bloque económico (rama de actividad y ocupación) no fueron aplicadas a toda la población, sino que se realizó una muestra para las ciudades de acuerdo a su cantidad de habitantes, dependiendo la magnitud de ésta del tamaño de las poblaciones. Así, en las ciudades más grandes se llegaron a relevar el 10% de los casos mientras que en las más pequeñas se indagó al total. En conjunto, para todo el país, aproximadamente el 50% de la población censada respondió a esas preguntas.

Dadas las demoras producidas antes del comienzo de la codificación, se decidió realizar una muestra sobre la ya tomada anteriormente. La idea era no prolongar más la publicación de información y acotar el tiempo de codificación¹⁰. La información publicada corresponde a la expansión del total codificado, que fue aproximadamente el 15% del total del universo de ocupados.

La codificación se realizó en forma manual, directamente sobre las cédulas censales. Esto trajo dos consecuencias:

- un sistema complicado y delicado de administración de las cédulas censales;
- la introducción de un sesgo y un error de cada codificador individual.

Para evitar el crecimiento de estos factores se efectuó un control de calidad, encontrándose el error total para la pregunta de ocupación dentro de los parámetros internacionales¹¹.

En el 2001 las condiciones tecnológicas y la dinámica de este trabajo fueron sustancialmente distintas. No sólo porque una gran parte de los casos serían codificados automáticamente, sino porque además al resto se los trataría de manera semi-manual, a través de un conjunto de codificadores que trabajarían cada uno con una PC conectada en red. Esto fue posible gracias a la decisión del INDEC de capturar con un scanner la totalidad de la información censal, incluyendo la de las preguntas abiertas¹², lo que permitió evitar la manipulación de las cédulas censales y agilizar mucho su administración y tratamiento¹³.

De todos modos restaban dos cuestiones a resolver, por un lado el manejo del volumen de casos a codificar -alrededor de 11 millones de ocupados- directamente relacionado con el tiempo y número de codificadores requeridos; y por el otro el sesgo y el error introducidos por los codificadores.

¹⁰ La codificación de ocupaciones comenzó en octubre de 1993 y finalizó en agosto de 1995, empleándose un tiempo ubicado dentro de los parámetros internacionales para las preguntas con estas características.

¹¹ El error fue del 12% aproximadamente. Ver *Desarrollo y evaluación del tratamiento de las ocupaciones en el CEN '91*, Documento de Trabajo, julio 1996, INDEC.

¹² En el Censo de 1991 sólo se había capturado la información de las preguntas cerradas.

¹³ Hay que tener en cuenta que en el Censo de 1991, el tiempo que les insumía a los codificadores la búsqueda de los casos a codificar entorpecía todo el proceso. Primero debían separar las cédulas censales en blanco correspondientes a los no ocupados, ya que las de los ocupados no podían diferenciarse de las de los otros miembros de su hogar y además cada cédula incluía muchas hojas entre las cuales había que encontrar la que finalmente contenía la información ocupacional a codificar. Con la codificación semi-manual se resuelve este problema ya que el codificador recibe automática e inmediatamente un nuevo caso cuando ha puesto y confirmado el código al caso que tenía en pantalla.

Para mejorar el tratamiento de estos problemas es que se decidió elaborar el SiCI. Con su implementación se logró codificar automática y rápidamente el 57,3% de los ocupados de todo el país. A su vez, dado que esta etapa fue realizada por un sistema informático a partir de los parámetros previamente establecidos por el equipo conceptual de codificación, todos los casos similares codificados automáticamente fueron resueltos de la misma manera, eliminando de este modo el sesgo que introducía cada codificador.

El proceso de codificación de ocupaciones en el 2001

Sintéticamente, el tratamiento general de la información a codificar constó de los siguientes pasos:

1. El equipo de coordinación general del SiCI transfería a la sede Tejedor del INDEC la base de datos de una provincia completa, con la información necesaria para la codificación ocupacional y los datos de identificación de los casos. El equipo de coordinación del ProMAEO ejecutaba sobre ella una serie de programas diseñados para la codificación informatizada, programas que estaban almacenados en una de sus PC.
2. El resultado de la aplicación de los programas de codificación informatizada era por un lado, una base de datos que incluía a todas las ocupaciones que habían sido codificadas automáticamente y sus respectivos códigos, y por otro, una serie de bases de datos que abarcaba a todos los casos que no habían sido codificados automáticamente (casos no codificados y casos embolsados) en un formato adecuado para ser ubicado en la red de codificación semi-manual. La cantidad de bases de datos de dicha serie dependía del número de ocupados de la provincia que no hubieran sido codificados automáticamente.
3. Las bases para la codificación semi-manual, luego de un previo acondicionamiento, eran puestas en la red y retiradas al ser completadas por los codificadores. Sobre las bases ya codificadas en forma semi-manual se realizaba un control de calidad, explicado en otra parte de este informe.
4. Las bases codificadas automáticamente eran sometidas a un proceso de control de calidad, tal como se explicará en el capítulo siguiente. De este proceso podían derivarse correcciones o agregados a los microprocesos de codificación por palabras clave o a su secuencia. Además, del análisis de toda la información (codificada automáticamente o no) podía surgir la elaboración de nuevos microprocesos, algunos de ellos para ser aplicados de allí en más a las siguientes provincias, otros específicos para la provincia en consideración¹⁴.
5. Una vez terminada la codificación de una provincia, se unían todas las bases de la codificación semi-manual a la de la codificación automática y se transferían al equipo de coordinación general del SiCI.

¹⁴ Si bien teóricamente era posible volver a ejecutar los microprocesos para las provincias anteriores, ello podía traer problemas de administración de las bases de datos, dado que los casos que no habían sido codificados automáticamente ya habían sido resueltos en la línea semi-manual. Por lo tanto, para evitar superposiciones o pérdida de casos se resolvió no volver a ejecutar los microprocesos en las provincias ya codificadas.

A efectos de una mejor exposición y comprensión, los resultados obtenidos se presentan por separado de acuerdo a la forma de codificación: automática y semi-manual.

La codificación automática

El objetivo principal del SiCI es la posibilidad de codificar automáticamente -es decir sin la participación de codificadores- la mayor cantidad de casos con el menor porcentaje de error y en un tiempo reducido.

Los resultados obtenidos superaron largamente las expectativas iniciales de este equipo. En una de las primeras pruebas realizada en el año 2000 con la información del censo experimental efectuado en la ciudad de Pergamino, se alcanzó el 35% de codificación automática. Tal como se presenta en el cuadro que sigue, en el Censo de Población 2001 se codificó automáticamente, para el total del país, el 57,3% de los casos, es decir alrededor de 6.280.000 ocupados.

Cuadro Nro. 1: Porcentajes de codificación automática y semi-manual sobre total de ocupados. Censo 2001¹⁵.

Provincia	Total ocupados	codificación automática %	codificación semi-manual %
Neuquén	152577	53.2	46.8
Misiones	259660	62.0	38.0
Cdad. de Bs.As.	1163404	<u>51.0</u>	49.0
Chaco	240425	62.2	37.8
Corrientes	235146	57.0	43.0
Jujuy	171508	58.0	42.0
Salta	278407	59.8	40.0
Mendoza	468021	57.1	42.9
Formosa	118527	61.5	38.5
Córdoba	1027479	59.8	40.2
La Pampa	110804	61.5	38.5
La Rioja	88700	52.8	47.2
Santa Fe	908934	57.5	42.5
Río Negro	169053	57.6	42.4
Chubut	136907	53.8	46.2
Tucumán	329956	58.0	42.0
24 partidos GBA	2424098	56.2	43.8
Catamarca	94180	55.9	44.1
Tierra del Fuego	38408	54.0	46.0
Resto pcia. Bs.As.	1647208	59.5	40.5
Stgo. Del Estero	188237	60.9	39.1
Santa Cruz	73486	<u>52.2</u>	47.8
Entre Ríos	335859	59.9	40.1
San Luis	114284	58.9	41.1
San Juan	174957	57.2	42.8
Total	10950225	57.3	42.7

¹⁵ Se presentan las provincias de acuerdo al orden cronológico en el que fueron codificadas.

Las diferencias provinciales en los porcentajes de codificación automática dependen de la diversidad ocupacional dentro de cada una de ellas. Los microprocesos se concentran básicamente en las ocupaciones más frecuentes y comunes a todas las jurisdicciones, quedando en general fuera aquellas con menor frecuencia y las muy específicas de cada región.

Un ejemplo en el que la variedad de ocupaciones es alta y por lo tanto hay una menor frecuencia de las más comunes, es la Ciudad de Buenos Aires que, por esta razón, concentró el valor relativo más bajo de codificación automática de todo el país. Asimismo provincias como Santa Cruz o Chubut, con un desarrollo importante de la industria petrolera, también presentaron una baja proporción de casos codificados automáticamente pues era difícil conocer de antemano la estructura ocupacional contenida en esa actividad y desarrollar procesos automáticos pertinentes. Contrariamente, en provincias donde esas ocupaciones comunes o más frecuentes -ya captadas por los microprocesos- representan una parte muy significativa de su estructura, el porcentaje codificado fue mayor, como por ejemplo ocurre con Misiones, Chaco y Formosa, entre otras.

El tiempo de procesamiento requerido para cada provincia fue muy bajo, variando entre dos horas, para las provincias con menos ocupados, y un par de días para las que concentraban una gran cantidad de aquéllos, como 24 Partidos del Gran Buenos Aires o Ciudad de Buenos Aires.

La codificación semi-manual

Como ya se ha señalado, los casos que no fueran codificados automáticamente, serían resueltos de manera semi-manual por un conjunto de codificadores que, entrenados adecuadamente, pudieran completar la totalidad de la tarea. El esquema de trabajo pensado para esta instancia era el siguiente:

1. Curso de capacitación de los codificadores.
2. Codificación semi-manual a través de una red de computadoras.
3. Capacitación permanente durante la codificación.

Cada curso de capacitación duró dos semanas. En la primera se introdujo a los codificadores en la temática ocupacional y en las características del clasificador. Se realizó una introducción conceptual general, se presentaron en forma secuenciada las dimensiones y los dígitos del clasificador y se efectuaron ejercicios a tal efecto. Durante la segunda semana, se continuó la ejercitación de casos pero ahora en la pantalla a emplear durante el proceso mismo de codificación, situación que sirvió tanto de práctica de codificación como de acercamiento al programa a utilizar. Se introdujo además en el uso de materiales de clasificación complementarios al código¹⁶.

Dado que la incorporación de los codificadores fue paulatina, se implementaron cuatro cursos de capacitación, tal como se consigna en la tabla que sigue:

¹⁶ Junto a los instrumentos de capacitación se diseñaron materiales complementarios al código que operaron como elemento de apoyo para la codificación. Ellos son: un conjunto de reglas operativas que brindaron criterios uniformes para resolver las dificultades que presentaba la información y un Manual de Capacitación. Para conocer más detalles puede consultarse INDEC, Manual de Capacitación de Ocupaciones, Versión 2001, pag. 87. (mimeo).

Fecha de los cursos	Cantidad de codificadores
1. 26 de agosto al 6 de septiembre de 2002	7 codificadores
2. 9 al 20 de diciembre de 2002	31 codificadores
3. 24 de febrero al 7 de marzo de 2003	18 codificadores (7 estuvieron en el proceso de codificación) ¹⁷
4. 2 al 12 de junio de 2003	10 codificadores

Una vez finalizados los cursos de capacitación, los codificadores comenzaban a clasificar las ocupaciones que no habían sido resueltas automáticamente. El proceso de codificación semi-manual funcionaba del siguiente modo:

1. Cada codificador recibía en una pantalla de PC un caso por vez, con toda la información necesaria (incluyendo las de las preguntas complementarias a la de ocupación) a la que le introducían el código ocupacional correspondiente.
2. La información le llegaba ordenada. Los administradores in-situ del sistema de codificación (miembros del ProMAEO) establecían los criterios para ello, de modo que los casos aparecieran ordenados por afinidad de las variables complementarias (casos de una misma categoría ocupacional, por ejemplo) y por el orden alfabético de las respuestas de los censados a las preguntas abiertas. Esto hacía que muchos casos iguales o similares se sucedieran sin interrupción y que el codificador no tuviera que volver a pensar más tarde una resolución que ya había encontrado (como era el caso con la codificación sobre las cédulas censales, cuyo único criterio de orden era logístico-administrativo).
3. El sistema incorporaba una tabla de validación de códigos ocupacionales. De esta manera, si el codificador ingresaba un código no válido el mismo era rechazado por el Sistema (esto por supuesto no impedía que un código válido pero incorrecto fuera adjudicado.) Dichas tablas de validación no funcionaron desde el inicio del proceso, sino hasta que el problema fue advertido y corregido.

La velocidad de la codificación semi-manual superó ampliamente a la de todas las experiencias anteriores de codificación manual e incluso a la de codificación común en pantalla realizada a veces por miembros del ProMAEO.

La codificación en red

Si bien la idea era trabajar desde el principio con una red interna para la codificación semi-manual, no fue posible hacerlo inmediatamente, ya que requirió algunos ajustes entre el programa de codificación y la red disponible en la sede Tejedor del INDEC. Por lo tanto, la primera provincia codificada allí (Misiones) y una parte de la segunda (Ciudad de Buenos Aires) requirió ser dividida en múltiples fragmentos que eran distribuidos y retirados uno a uno en las computadoras de los codificadores, tarea que estuvo a cargo de los administradores in-situ del sistema de codificación. Poco después fue solucionado el inconveniente mencionado y todos los codificadores obtenían sus casos automáticamente desde una base única ubicada en una red interna. Esto aceleró la codificación y simplificó enormemente la administración de las bases de datos, reduciendo el riesgo de errores.

¹⁷ Los 11 restantes se capacitaron para asistir a otros programas del INDEC.

Además del acondicionamiento alfabético y el orden general que se les daba a las bases, se las fraccionaba cada 60.000 casos, lo que permitía una más fácil manipulación: con este tamaño era posible realizar chequeos rápidos con el programa Excel cuando era necesario.

Precisamente el acondicionamiento alfabético fue el que permitió eliminar, para la codificación semi-manual, los llamados *casos embolsados*. Tal como se apuntó en el capítulo 1, estos eran casos que el sistema agrupaba por palabras clave comunes que no permitían su codificación total pero sí su agrupamiento temático. La idea inicial era especializar a distintos grupos de codificadores en distintas bolsas para así concentrar esfuerzos en casos similares. La presentación ordenada alfabéticamente de los casos a codificar convirtió en superfluas a las bolsas. Por lo tanto, no tenía sentido separar y especializar a grupos de codificadores. Los casos embolsados fueron tratados exactamente igual que los casos no codificados automáticamente. En un principio, los códigos genéricos o de bolsa fueron reemplazados en la línea semi-manual por los codificadores, quienes recibieron la instrucción de ignorarlos y codificar los casos igual que los otros que recibían. Más adelante, el equipo de administración de bases de datos del ProMAEO borró el código genérico antes de que el caso llegara a la línea de codificación semi-manual.

Un dato importante para la codificación semi-manual es que, dado que el clasificador utilizado supone códigos distintos para categorías ocupacionales distintas, se diseñó una manera de abordaje en la cual el codificador seleccionaba una categoría ocupacional y a partir de allí sólo codificaba los casos en ella incluidos hasta que decidiera trabajar con otra.

Finalmente, otra de las instancias del proceso de codificación semi-manual lo constituye la capacitación permanente. El objetivo de esta tarea era mejorar la calidad de la información codificada restringiendo el margen de error a partir de la homogeneización de los criterios clasificatorios y recapacitando a los codificadores en función del análisis de los errores detectados en las bases de control de calidad. El equipo del ProMAEO, una vez finalizado el control de calidad de lo codificado en forma semi-manual, sistematizaba los errores a los efectos de volver a capacitar a los codificadores en esos temas.

Una vez codificadas todas las bases correspondientes a una provincia y terminado su control de calidad, el equipo de administración in-situ del ProMAEO volvía a unirlas y junto con la base resultante de la codificación automática la remitía al administrador central del SiCI.

Indicadores del funcionamiento de la codificación ocupacional

Desde el comienzo se planeó trabajar con un conjunto de indicadores que dieran cuenta del funcionamiento general del proceso de codificación semi-manual. En procesos de esta envergadura, que involucran la administración de un volumen importante de información y con un plazo de realización definido, se hace necesario contar con buenos indicadores cuantitativos que muestren la evolución del trabajo, los avances o retrocesos, la identificación de problemas si los hubiera y la evaluación de los ritmos laborales que permitan cumplir la tarea en el tiempo acordado. Además, se consideró la importancia de dejar registro de este tipo de experiencia que sirviera como parámetro real para futuras estimaciones de operativos de esta naturaleza.

Se decidió que la unidad de tiempo que mejor ordenaba todo el trabajo era el mes, pues se consideró que ese lapso era el apropiado para medir y observar la evolución de todos los indicadores.

Los indicadores con los que se trabajó fueron cuatro:

la productividad media diaria por codificador: el promedio diario de lo codificado por una persona en un mes, considerando la cantidad de jornadas reales en las que estuvo presente;

la producción media diaria: la totalidad de lo codificado por el conjunto de los codificadores en un mes dividida por la cantidad de días hábiles de ese mes;

el promedio diario de codificadores: la suma de codificadores presentes por día dividida por la cantidad de días hábiles del mes;

la producción mensual: la suma del total codificado en un mes.

La información de base con la que después se construían estos indicadores era ingresada diariamente en planillas de cálculo por el equipo del ProMAEO. El resultado para el lapso que duró la codificación es el que se presenta en el cuadro siguiente.

Cuadro Nro. 2: Indicadores seleccionados del proceso de codificación. Censo 2001.

Mes	Producción mensual	Productividad media diaria por codificador	Producción media diaria	Promedio diario de codificadores
Ene-03*	124410	301	8886	29
Febrero	159024	372	7951	21
Marzo	173334	297	8667	30
Abril	240327	335	12016	35
Mayo	321198	433	15295	35
Junio	409915	463	20495	44
Julio	503875	511	22903	45
Agosto	487026	530	24351	46
Septiembre	601391	595	27366	46
Octubre	620198	639	28191	44
Noviembre	450711	525	22536	43
Diciembre	316306	418	17573	42
Ene-04*	202173	492	20217	41

(*) La codificación comenzó a mediados de enero del 2003 y finalizó a mediados de enero del 2004.

Como puede apreciarse, en los tres primeros meses se observan los valores más bajos de todas las series. Por tratarse del comienzo, obviamente no existía ni la experiencia ni la seguridad en el manejo del clasificador, como tampoco la familiaridad con las respuestas censales. Asimismo, esos fueron los meses en los que una parte considerable de los codificadores hicieron uso de su licencia ordinaria. Ya a partir de abril el número promedio de codificadores aumentó y desde junio se instaló en valores similares hasta el fin del trabajo.

Los tres indicadores de producción aumentan progresivamente hasta el mes de octubre. En los últimos dos meses y medio¹⁸ decrecen como es usual en procedimientos repetitivos extendidos en el tiempo. También sucedió así en el Censo 1991.

¹⁸ En enero del 2004 se codificó durante los primeros quince días. En las siguientes dos semanas se terminaron las tareas de control de calidad.

En el capítulo siguiente se describirá cuál fue el método de estimación del error para conocer la calidad de los resultados obtenidos.

Después de haber leído este capítulo se debe de comprender que la estimación del error es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

Después de haber leído este capítulo se debe de comprender que la estimación del error es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

Después de haber leído este capítulo se debe de comprender que la estimación del error es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

Después de haber leído este capítulo se debe de comprender que la estimación del error es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

El estudio de calidad de la producción industrial

El estudio de la calidad de la producción industrial es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

El estudio de la calidad de la producción industrial es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

El estudio de la calidad de la producción industrial es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

El estudio de la calidad de la producción industrial es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

1. Este estudio de la calidad de la producción industrial es un proceso que se realiza en el momento de la recolección de los datos y que se debe de tener en cuenta desde el momento de la recolección de los datos hasta el momento de la estimación del error.

3. Control de calidad de la información ocupacional codificada

Tiempo antes de iniciarse las tareas de codificación censal se establecieron, entre los equipos intervinientes¹⁹, los objetivos que el control de calidad de este proceso debía perseguir. Se acordó que serían fundamentalmente tres: a) conocer el error por cada base codificada²⁰ y en consecuencia el error general de la codificación, aceptándose hasta un 15%; b) utilizar los resultados obtenidos en la capacitación de los codificadores y c) corregir y agregar microprocesos para la codificación automatizada.

Sobre el segundo punto se convino que, como no se contaba con el tiempo suficiente para procesar muestras que permitieran conocer además del error general, su distribución precisa por persona, la capacitación sería para todos por igual a partir del error general de cada base.

Dado que el procesamiento de la información ocupacional fue dividido en dos instancias -una automática y otra semi-manual-, también debió desdoblarse su control de calidad.

Para tal fin se establecieron pautas metodológicas precisas en el tratamiento de las dos instancias que se describirán a continuación.

El control de calidad de la codificación automática

De acuerdo a lo sugerido por la Dirección de Metodología Estadística, sobre cada base de datos resultante de la codificación automática se tomaba una muestra de alrededor de 2400 casos por provincia. Previamente a la extracción de la muestra, se ordenaban los casos según el microproceso que había permitido la codificación y según el código adjudicado; luego se extraía una muestra sistemática. Esta era analizada por miembros del equipo conceptual del ProMAEO, quienes determinaban qué casos eran erróneos.

El objetivo del control de calidad, además de permitir contar con una estimación del error de codificación, fue evaluar el funcionamiento de los microprocesos. De esta manera se fueron introduciendo correcciones y mejoras en ellos y en su secuencia.

En algunos casos se observaban errores fácilmente subsanables. Entonces se corregían los microprocesos en cuestión y a continuación se reiteraba la aplicación de los programas de codificación automática. Dado que el tiempo de procesamiento de éstos era mucho menor al de la codificación en la línea semi-manual, era posible adelantarse y realizar las correcciones antes de que fuera necesario entregar los casos sin codificar a esta línea. De todas maneras, por razones de administración de las bases de datos no era posible procesar más que la provincia inmediatamente siguiente a la que estaba siendo sometida a la codificación semi-manual.

En los casos de 24 partidos del Gran Buenos Aires y Resto de la Provincia de Buenos Aires, se implementó una segunda instancia de microprocesos. Dada la gran cantidad de ocupados en estas jurisdicciones, una vez terminado el procesamiento automático se decidió analizar los casos que no fueron codificados, con el objetivo de

¹⁹ Metodología Estadística y el SiNN.

²⁰ El error provincial que se presentará en este apartado es el resultado del control de todas las bases de una misma provincia.

estimar si era factible agregar nuevos microprocesos que aumentaran el porcentaje de lo codificado. El análisis confirmó que era posible, sumándose siete nuevos microprocesos que no sólo aumentaron en casi tres puntos porcentuales el total de lo codificado, sino que sirvieron para las provincias siguientes.

El control de calidad de la codificación semi-manual

El método utilizado en esta instancia variaba respecto del empleado en la anterior. Aquí el control se realizaba a través del procedimiento de pareo de la información que consiste en una doble verificación externa de lo codificado realizada por dos codificadores de la misma línea de codificación²¹.

Cada base de 60.000 casos terminada se ordenaba por codificador y por código asignado. De allí se extraía una muestra sistemática de 1200 casos de la cual se hacían dos copias para los verificadores, quienes codificaban los casos "a ciegas" y en las mismas condiciones que en la línea de codificación. Es decir que diariamente se sacaba de la línea de codificación a dos personas que pasaban a desarrollar la función de verificadores, quienes codificaban la base de control de calidad desconociendo el código asignado por el codificador y por el otro verificador. Seguidamente, el equipo del ProMAEO armaba una base única con los tres códigos y analizaba las distintas situaciones:

- a) si los códigos asignados por el codificador y los dos verificadores coincidían, el caso era considerado correcto;
- b) si coincidía el código asignado por el codificador y sólo uno de los verificadores el caso era correcto;
- c) si coincidían los códigos asignados por los dos verificadores pero no con el asignado por el codificador el caso era considerado incorrecto; y
- d) si los tres códigos eran distintos, el equipo de coordinación resolvía si la asignación dada por el codificador de línea era o no correcta.

En el apartado que sigue se presenta la estimación del error total y por provincia resultante tanto de la instancia automática como de la semi-manual.

Estimación general del error en la codificación de ocupaciones

El análisis en este apartado se refiere al error en su conjunto, sin discriminar por dimensión²². Se estableció que el parámetro que se aceptaría para el total de la información sin necesidad de recodificar, sería del 15%. Dado que la complejidad era mayor en los casos que se resolverían de manera semi-manual, éstos eran susceptibles de contener mayor error.

Efectivamente, tal como se observa en el cuadro 3, el porcentaje de error en una y otra instancia, confirmó tal diferencia.

²¹ Es el mismo método que se utilizó para controlar la calidad de lo codificado en el Censo de 1991.

²² Cabe aclarar que este error total es el error límite que se encontraría si se analizara para cada una de las cuatro dimensiones ocupacionales tomadas por separado. Esto es así ya que se considera errónea la codificación de un caso que contenga error aunque sea en una sola dimensión. Por lo tanto, los casos erróneos lo son en buena medida para alguna de las dimensiones y no para todas, por lo cual el error total supera al de cada dimensión en particular.

Cuadro Nro. 3: Error de codificación automática y semi-manual por provincia²³.
Censo 2001.

Provincia	Total ocupados	error codificación automática %	error codificación semi-manual %	error total %
Neuquén	152577	6.5	23.0	14.2
Misiones	259660	5.3	17.7	10.0
Cdad. de Bs.As.	1163404	6.4	21.2	13.7
Chaco	240425	6.2	22.9	12.5
Corrientes	235146	6.4	19.9	12.2
Jujuy	171508	5.4	17.4	10.4
Salta	278407	8.0	18.6	12.2
Mendoza	468021	6.5	19.3	12.0
Formosa	118527	5.5	17.5	10.1
Córdoba	1027479	5.7	18.9	11.0
La Pampa	110804	8.7	20.6	13.3
La Rioja	88700	6.6	18.1	12.0
Santa Fe	908934	5.7	19.4	11.5
Rio Negro	169053	4.4	19.8	10.9
Chubut	136907	4.8	23.1	13.3
Tucumán	329956	3.9	18.6	10.1
24 partidos GBA	2424098	6.0	18.2	11.3
Catamarca	94180	4.5	18.3	10.6
Tierra del Fuego	38408	7.2	21.3	13.7
Resto de pcla. Bs.As.	1647208	5.7	18.6	10.9
Stgo. Del Estero	188237	6.6	17.8	11.0
Santa Cruz	73486	6.7	20.3	13.2
Entre Ríos	335859	4.3	19.6	10.4
San Luis	114284	4.8	15.9	9.4
San Juan	174957	5.9	19.3	11.6
Total	10950225	5.9	19.2	11.6

El nivel de error total de la codificación ha sido similar al registrado en 1991. La lectura del cuadro 3 denota que los valores del error de la codificación semi-manual conforman una distribución bastante homogénea con extremos que se sitúan en el 15.9% para la provincia de San Luis y el 23.1% para Chubut.

Hay varios factores que hacen que el error sea mayor en la codificación semi-manual que en la automática. En primer lugar, la propia lógica del diseño de codificación hace que los casos más sencillos de codificar y de mayor presencia cuantitativa sean codificados automáticamente, por lo cual los casos más complejos y variados pasan a la línea semi-manual. Esto conduce a que el error en esta línea sea no sólo mayor que en la automática, sino también que en la codificación del Censo 91, ya que se ha producido una concentración de casos difíciles y variados en la línea semi-manual. En el censo anterior, el error manual se veía atenuado por una gran cantidad de casos sencillos y de alta frecuencia que en el censo actual no llegaron a manos de los codificadores.

Por otra parte, dadas las limitaciones de tiempo con que se requería terminar la codificación (un año, la mitad que para 1991) se optó por no realizar un control de calidad en donde estuvieran representados con significación estadística los codificadores individuales, sino todo el conjunto. La información individual hubiera sido

²³ El orden de presentación de las provincias en este cuadro es el mismo en el que fueron codificadas.

relevante para diseñar una capacitación más personalizada, pero realizar un control de estas características hubiese exigido una muestra mucho mayor y una cantidad más grande de codificadores realizando tareas de control de calidad en detrimento de su participación en la línea de codificación, retrasándose así algunos meses la finalización del proceso. También hubiese significado la asignación permanente de miembros del equipo de coordinación de la codificación para el seguimiento y análisis de dicha información, recargando más allá de lo posible las tareas del equipo.

Por último, hay que tener en cuenta que se ha codificado semi-manualmente dos veces y media y en la mitad de tiempo la cantidad de casos del Censo de 1991. Esto supuso una aceleración del proceso de codificación que fue bastante mayor a lo esperado, por lo cual no hubo posibilidad, en medio del proceso, de planificar e implementar otras medidas que atenuaran el nivel de error.

A partir de esta experiencia se concluye que para próximos operativos es necesario planificar los recursos materiales y humanos que posibiliten llevar adelante un procedimiento de control de calidad que permita estimar el error por codificador y así diseñar una política de capacitación más específica y personalizada.

4. Codificabilidad y tratamiento de la información

Codificabilidad de la información ocupacional censal

La posibilidad de que la información relevada en campo pueda transformarse en un grupo ocupacional específico del CNO, se ve afectada por las características del instrumento de recolección, por la calidad del relevamiento y por la precisión del clasificador. En cualquier operativo a gran escala como el censal, es esperable que haya un porcentaje de casos de difícil resolución o incodificables. Cuanto más alto sea ese porcentaje, menor es la calidad de los resultados de codificación. En este sentido, la codificabilidad de la información es un indicador que permite evaluar, entre otros aspectos, el resultado de la aplicación del instrumento clasificatorio a la información censal.

Al proceso de codificación llegaban todos los casos en los cuales el censado se reconocía como ocupado en la pregunta de condición de actividad así como también aquellos casos en los cuales, independientemente de la declaración sobre la condición de actividad, existía texto en las preguntas ocupacionales. De esta manera el análisis de la codificabilidad de la información ocupacional censal permite –entre otras cosas– distinguir distintas situaciones cuyo tratamiento permitiría corregir errores en la condición de actividad y discriminar situaciones diferentes al interior del universo de ocupados.

En función de las respuestas a las preguntas sobre el nombre de la ocupación y la descripción de las tareas, pueden señalarse las siguientes posibilidades en relación a la codificabilidad.

- **casos con información completa:** cuando dada la información recolectada en campo, es posible reconocer el valor de las cuatro dimensiones del CNO y por lo tanto la ocupación a codificar puede clasificarse en un grupo ocupacional del CNO;
- **casos con información parcial:** cuando en la ocupación a codificar se desconoce el valor de al menos una de las cuatro dimensiones, pero no de todas, por ejemplo: 99313 (se desconoce el *carácter*), 40319 (se desconoce la *calificación*);
- **casos con información insuficiente:** cuando la información proveniente de campo no alcanza para clasificar a la ocupación en ningún grupo ocupacional del CNO, es decir no permite reconocer las tareas concretas que definen a la ocupación a codificar, por ejemplo: "empleado", "tareas generales". En esos casos, se asigna el código 99997. Claramente, en estos casos, las dificultades provienen del relevamiento;
- **casos con información en blanco:** son los casos en los que no se registra información alguna en las preguntas de ocupación. En estos casos, se asigna el código 99999;
- **información que no corresponde codificar:** son los casos en los cuales, si bien aparece información en las preguntas sobre ocupación y tarea, la misma refiere a la definición censal de no ocupados. Al igual que la situación descrita anteriormente, éstas son dificultades inherentes al relevamiento. Los casos encontrados referían a:

información de población inactiva: aquellos casos en los cuales por las respuestas a las preguntas de ocupación y tarea se deduce que no se trata de trabajadores activos (amas de casa, estudiantes o voluntarios) o no integran el universo de la población económicamente activa de la Argentina (trabajadores de embajadas extranjeras en el país). En esos casos se les asignaba el código 99998;

información de población desocupada: son los casos en los que en las preguntas de ocupación el censo se declara como desocupado. El código correspondiente es el 99996.

Los cuadros que se presentan a continuación ordenan los resultados de codificabilidad de acuerdo a como aparecieron en el Censo de 1991, para comparar su comportamiento en ambos censos. A tal efecto, los porcentajes que aparecen en el cuadro 4 corresponden al cociente entre la suma de los casos con información insuficiente (99997) y los que no corresponde codificar (99998) respecto al conjunto de la información ocupacional codificada, para cada provincia y para el total del país.

Cuadro Nro. 4: Casos con información insuficiente y que no corresponde codificar. Provincias y total país, 1991-2001²⁴.

Provincia	Casos con información insuficiente y que no corresponde	
	1991	2001
	%	%
Cdad. de Bs.As.	7.2	2.2
Catamarca	9.2	3.6
Córdoba	7.7	2.3
Corrientes	8.7	4.2
Chaco	8.1	2.8
Chubut	8.3	2.9
Entre Ríos	8.7	3.7
Formosa	9.8	3.8
24 partidos GBA	7.6	4.3
Jujuy	6.9	1.7
La Pampa	7.7	2.0
La Rioja	10.0	2.6
Mendoza	6.4	1.9
Misiones	12.3	5.4
Resto de pcla. Bs.As.	8.0	3.3
Neuquén	8.9	3.8
Río Negro	9.6	3.1
Salta	7.6	2.0
San Juan	7.9	2.5
San Luis	10.2	3.1
Santa Cruz	15.6	3.2
Santa Fe	7.4	2.6
Sgo.del Estero	8.5	3.8
T. del Fuego	10.8	2.9
Tucumán	8.3	2.6
Total	8.0	3.2

²⁴ Datos provisorios. Por asignación pueden modificarse los valores.

La disminución en la proporción de estos casos en el 2001, tanto para el total del país como para cada una de las provincias, denotan un mejoramiento tanto de la recolección de la información como de la aplicación de la nueva versión del clasificador, además de codificadores mejor entrenados. No se ha avanzado, para este informe, en una evaluación particular de cada uno de estos componentes, por lo que no se puede aseverar que alguno incida más que otro.

También la proporción de casos con información parcial disminuyó significativamente, como se observa en el cuadro 5, en especial los casos con *carácter* ignorado. Al desagregar el clasificador en más grupos, en lo que a esta dimensión respecta, y mejorar sensiblemente las reglas operativas, se logró aumentar la codificabilidad de esta dimensión.

Cuadro Nro. 5: Casos con información parcial. Total país, 1991-2001.

Total país	1991 %	2001 %
Calificación Ignorada	4.0	2.2
Carácter Ignorado	3.1	0.1
Total	7.0	2.4

Algunas dificultades que presentaba la información a codificar

Las características de la información proveniente de este operativo censal obligó a tomar decisiones en el propio proceso de codificación para resolver las inconsistencias detectadas entre las respuestas a las preguntas de condición de actividad, categoría ocupacional y las declaraciones ocupacionales.

Además de los problemas ya presentados relacionados con la posibilidad de codificar las declaraciones ocupacionales de los censados que básicamente se debían a la insuficiencia de la información o a la detección de declaraciones de no ocupados, durante el proceso de codificación se encontraron otras dificultades que implicaron la toma de decisiones no previstas inicialmente. Algunos de los problemas que se reseñarán han sido detectados y resueltos a partir de reglas de consistencia y patrones de asignación; otros fueron corregidos en el proceso mismo de codificación y sus soluciones incorporadas a las reglas de codificación. En todos los casos, las decisiones fueron tomadas en forma conjunta por este equipo de trabajo y el de Diseño Conceptual.

Entre las principales dificultades encontradas figuran los casos de no concordancia entre la categoría ocupacional y la descripción de las tareas realizadas por el censado. En ese sentido sobresalen quienes se declaraban trabajadores por cuenta propia pero por las respuestas en las preguntas de rama de actividad y nombre de la ocupación, se desprendía que no lo eran. Las situaciones y soluciones encontradas han sido las siguientes:

- a) Personal militar de las Fuerzas Armadas o personal policial que aparecen como trabajadores por cuenta propia. Se decidió no cambiar la categoría ocupacional, por lo tanto en esos casos fue asignado el código de información insuficiente (99997).

- b) Legisladores, jueces y funcionarios del Poder Ejecutivo que aparecen como trabajadores por cuenta propia. Estos fueron codificados en su correspondiente grupo de carácter –Ocupaciones de Dirección- y por medio de las pautas de consistencia fue posible cambiarles la categoría ocupacional y considerarlos como asalariados.
- c) Trabajadores por cuenta propia que se declararon como jefes. La propia definición de esta categoría exime a los cuenta propia de tener personal a cargo y por lo tanto de ocupar ese valor en la dimensión jerarquía. En estos casos, fue respetada la categoría ocupacional y por lo tanto considerados como trabajadores por cuenta propia sin personal a cargo.
- d) Las empleadas domésticas que se clasificaron como cuenta propias: En el operativo censal, a pesar de que éstas se declaraban como trabajadores por cuenta propia, eran codificadas como asalariadas y por lo tanto a partir de reglas de consistencia se les cambiaba la categoría ocupacional. En ese sentido se decidió respetar los criterios seguidos en la EPH a partir de su reformulación.

De los casos expuestos surge que estos problemas están relacionados con la forma de captación de la variable categoría ocupacional, en la que el censado se autodefine eligiendo una de las opciones. Si bien se realizaron pruebas piloto en las cuales específicamente se probó un conjunto de indicadores para reconstruir la variable categoría ocupacional, se decidió no contemplar esa posibilidad y mantener las preguntas de autodefinición. A partir de la experiencia obtenida en el último censo, y a la luz de los problemas detectados más arriba, probablemente habría que reconsiderar esta decisión.

5. Conclusiones

Luego de varios años de trabajo, el balance entre las metas propuestas y los resultados obtenidos es positivo. Una parte importante de lo señalado en las conclusiones de la experiencia del censo anterior²⁵, en el sentido de los pasos a seguir para superar y mejorar las distintas etapas de producción de la información económica, pudo ser retomado y puesto en práctica.

En grandes líneas y para resumir lo apuntado en los apartados correspondientes de este informe, las mejoras estuvieron localizadas en:

- el cambio en la formulación de las preguntas en la cédula censal, relevando más y mejores respuestas ocupacionales;
- una nueva versión del clasificador de ocupaciones, a cinco dígitos, que permitió una mayor precisión en la clasificación de varios grupos de ocupaciones;
- el diseño, desarrollo, puesta a punto y aplicación de un Sistema de Codificación Informatizado²⁶, que creó las condiciones de codificación de la *totalidad* de las respuestas dadas por la población en un lapso mucho menor²⁷ a experiencias pasadas, en las que, como en 1991, sólo se codificó una muestra;
- la puesta en marcha de un proceso de codificación semi-manual a través de una red de computadoras, que permitió incrementar la productividad por codificador y la producción diaria a niveles inesperados en las estimaciones previas;
- un menor porcentaje de información incodificable o con codificación parcial, lo que implica mejores datos.

En cuanto al error, si bien en lo codificado automáticamente se bajó a la mitad del error general, en la instancia semi-manual el error se duplicó, por las razones que ya fueron expuestas en el capítulo 3. La experiencia ha demostrado que es necesario contar con esta información discriminada por codificador, a efectos de establecer una capacitación personalizada. Para ello, las muestras deberán ser más grandes y por tanto hará falta disponer de un número mayor de personas que trabajen en la verificación, si se quiere mantener un ritmo simultáneo con la información que se codifica en línea. Además, la posibilidad de un ulterior aumento del porcentaje de codificación automática, disminuyendo la cantidad de casos a ser resueltos semi-manualmente, creará mejores condiciones para aquella tarea.

Con respecto al relevamiento general y a partir de los problemas señalados respecto de la captación de la categoría ocupacional, sería interesante volver a contemplar la posibilidad de medir dicha variable no a partir de la autodefinición del censado sino a través de distintos indicadores que, reconstituidos en gabinete, brindarían una medición más válida. De esta manera además, se mejoraría el tratamiento de las variables económicas.

²⁵ Ver op. cit. Conclusiones, pág. 51.

²⁶ La aplicación del SiCI, por otra parte, trasciende al Censo Nacional de Población. Con ligeras adaptaciones se puede utilizar en otros programas y áreas. Por ejemplo, ya se está utilizando en la codificación de la Encuesta Permanente de Hogares y se aplicará en la próxima Encuesta Nacional de Gastos de los Hogares, y puede serlo en cualquier programa que obtenga una respuesta abierta sobre rama de actividad, ocupación, etc.

²⁷ Con respecto al Censo de 1991, se codificaron aproximadamente 6 veces más casos en la mitad de tiempo.

Finalmente el SiCI ha probado ser un instrumento eficiente para lograr una codificación rápida y de calidad de las variables económicas en el Censo 2001. A su vez, actualmente se está iniciando su implementación en la EPH y se prevé su inclusión en otros relevamientos del Instituto. En ese sentido, sería importante no sólo promover su uso al interior del INDEC sino también en otros organismos del Estado. Para ello sería necesario avanzar en el registro del SiCI a los efectos de que el INDEC resguarde la propiedad intelectual de este producto.

Cabe destacar que el éxito del operativo no hubiese sido posible sin la colaboración de las distintas Direcciones que aportaron su capacidad técnica y profesional, como así también del grupo de codificadores que hizo posible superar las expectativas. En este sentido, sin duda, haciendo un análisis comparativo y cualitativo con el Censo '91 se ha mejorado significativamente.

6. Equipo de codificadores

- 1) Abdala Mayra
- 3) Arias Adriana
- 4) Bardón Rodolfo
- 5) Caram Silvia
- 6) Cortes Alberto
- 7) Delfino Néstor
- 8) De Marco Sergio
- 9) Encalada Mercedes
- 10) Enatarriaga Héctor
- 11) Fernández Gallo Sonia
- 12) Glaz Omar
- 13) Iamurri Luisa
- 14) Kelsch Karina
- 15) Loíacono Gustavo
- 16) Lomban Jorge
- 17) López Alejandra
- 18) Mendoza Horacio
- 19) Mojardin Raúl
- 20) Moyano César
- 21) Ochoa Inés
- 22) Orbe Sandra
- 23) Paez Lilian
- 24) Parada Nora
- 25) Palacio Cecilia
- 26) Prieto Graciela
- 27) Ramírez Antonia
- 28) Ramos Claudia
- 29) Rivas Yolanda
- 30) Rodríguez Leandro
- 31) Rust Juan
- 32) Santana Beatriz
- 33) Sánchez Jorge
- 34) Sarmiento Alicia
- 35) Vicenti Norma
- 36) Zapata, Zulma.
- 37) Furer, Javier.
- 38) Quiñones, María.
- 39) Romano Muñoz, Carlos.
- 40) Espasandín, Javier.
- 43) Barrios Esther
- 44) González, Leonor
- 45) Parolaro, Myriam
- 46) Botta, Silvio
- 47) Carrizo, Carina
- 48) Cevallos, Silvia
- 49) Cowes, María
- 50) Dias Coelho, Laura
- 51) Esses, Fabiana
- 52) Movia, Helios
- 53) Muratone, María
- 54) Oliari, Flavia