

Censo Nacional de Población, Hogares y Viviendas 2022

Codificación de las preguntas abiertas

Febrero de 2026

Censo Nacional de Población, Hogares y Viviendas 2022

Codificación de las preguntas abiertas

Febrero de 2026

Instituto Nacional de Estadística y Censos (INDEC)

Dirección: Pedro Ignacio Lines

Dirección Nacional de Difusión y Comunicación: María Silvina Viazzi

Dirección Nacional de Metodología e Infraestructura Estadística: Gerardo Mitas

Coordinación de Clasificadores y Nomenclaturas: Claudio La Rocca

Los informes metodológicos que conforman esta publicación fueron producidos por:

Equipo de País de origen: Marcelo Esses

Equipo de Pueblo indígena u originario: Anastasia Cenóz, Javier Domench, Claudio La Rocca

Equipo de Rama de actividad económica: Dolores González, Marcelo Esses

Procesamiento informático: Gerardo Graniel, Estefanía Martone

Coordinación de Producción Gráfica y Editorial: Marcelo Costanzo

Diseño y diagramación: Claudio Las Heras

Revisión y corrección: Horacio Barisani, Juliana De Glée y Martín Van Houtte

Agradecimientos

Al equipo de la Dirección Nacional de Estadísticas Sociales y de Población; a la Dirección General de Tecnologías de la Información y las Comunicaciones; a la Dirección de Logística Operativa, entre quienes se quiere hacer especial mención a Norberto Sposari. Al Instituto Nacional de Asuntos Indígenas (INAI) y al equipo de personas expertas referentes de los distintos pueblos indígenas, quienes aportaron su conocimiento e idoneidad en la tarea de codificación. Al equipo de la Secretaría de Extensión Universitaria de la Universidad Nacional de la Matanza (UNLaM): Mario Vilaltella, Gabriela Giménez y Leandro García.

ISBN 978-950-896-713-8

Instituto Nacional de Estadística y Censos (INDEC)

Censo nacional de población, hogares y viviendas 2022 : codificación de las preguntas abiertas. - 1a ed. - Ciudad Autónoma de Buenos Aires : Instituto Nacional de Estadística y Censos - INDEC, 2026.

Libro digital, PDF - (Censo nacional de población hogares y viviendas 2022 ; 15)

Archivo Digital: descarga y online

ISBN 978-950-896-713-8

1. Estadísticas. 2. Censos. 3. Censos Nacionales.

CDD 352.75



Libro de edición argentina

Buenos Aires, febrero de 2026

Publicaciones del INDEC

Las publicaciones editadas por el Instituto Nacional de Estadística y Censos están disponibles en www.indec.gob.ar y en el Centro Estadístico de Servicios, ubicado en Av. Presidente Julio A. Roca 609 C1067ABB, Ciudad Autónoma de Buenos Aires, Argentina. También pueden solicitarse al teléfono +54 11 51031-4632 en el horario de atención al público de 9:30 a 16:00.

Correo electrónico: ces@indec.gob.ar

Índice

1. Introducción.....	4
1.1 El Sistema de Codificación Informatizada (SiCI).....	4
2. Codificación de País de nacimiento	6
2.1 Instrumento de clasificación utilizado	6
2.2 Adaptación del Sistema de Codificación Informatizada	6
2.2.1 Relevamiento y codificación	6
2.2.1.1 Censo digital.....	7
2.2.1.2 Censo por barrido territorial	7
2.2.2 Revisión y cierre de las bases de datos.....	9
Anexo 1: Código de países	10
3. Codificación de Pueblo indígena u originario	12
3.1 Fases previas a la codificación	12
3.1.1 La elaboración del nomenclador de pueblos indígenas (NPI)	12
3.1.2 Decisiones metodológicas.....	13
3.2 El relevamiento: la captación de pueblos indígenas en las dos modalidades del censo	13
3.3 La codificación de la pregunta de pueblos indígenas.....	14
3.3.1 Codificación del relevamiento digital	14
3.3.1.1 Control de calidad del relevamiento digital	15
3.3.2 Codificación del relevamiento por barrido territorial.....	16
3.3.2.1 Condiciones de origen de los datos a codificar. Lectura de las respuestas.....	18
3.3.2.2 Codificación del relevamiento por barrido territorial en su modalidad automática	18
3.3.2.3 Codificación del relevamiento por barrido territorial en su modalidad asistida.....	18
3.3.2.4 Control de calidad del relevamiento por barrido territorial	19
3.4 Resultados del proceso de codificación.....	20
3.5 Conclusiones.....	22
Anexo 2. Nomenclador de pueblos indígenas	24
Pautas de codificación	24
4. Codificación de Rama de actividad económica.....	26
4.1 Fases previas al proceso de codificación	26
4.1.1 Consideraciones preliminares.....	26
4.1.2 Instrumentos para la clasificación y codificación	28
4.1.3 El Sistema de Codificación Informatizada y la codificación asistida. Su preparación.....	28
4.2 Proceso de codificación censal	28
4.2.1 Organización de la capacitación del equipo de codificación.....	29
4.2.1.1 La capacitación	29
4.2.2 Secuencia general en la codificación de la rama de actividad	30
4.2.2.1 Codificación de la pregunta 34.....	30
4.2.2.2 Tratamiento de la pregunta 33.....	30
4.2.2.3 Entrega de la información para cierre y consistencias	31
4.2.3 Estrategias y decisiones metodológicas	31
4.3 Resultados del proceso de codificación	32
4.4 Indicadores de seguimiento de la codificación asistida.....	37
4.5 El control de calidad	38
4.5.1 Estrategias de control de calidad	38
4.5.2 Acciones implementadas para la reducción del error en la codificación asistida	39
4.5.3 Niveles de error estimados.....	39
4.5.4 Fuentes de error identificadas	41
4.6 Conclusiones.....	41
Anexo 3. Diseño del formulario del bloque económico en el Censo 2001	43
Anexo 4. Proceso de codificación de la rama de actividad en el Censo 2022	44
Diagrama del proceso de codificación de rama de actividad en el Censo 2022	44
Anexo 5. Esquema de la secuencia general de pasos en la codificación de la rama de actividad	45
Anexo 6. Tabla de conversión y correspondencia de los valores consignados en la pregunta 34 (CODACT) a letra ClaNAE	46
Anexo 7. Tabla de conversión y correspondencia de letra ClaNAE a la pregunta 33 corregida	47
Anexo 8. Evaluación del proceso de codificación por parte de las codificadoras y los codificadores.....	48
5. Equipos de trabajo para la codificación	51
6. Bibliografía.....	52
Siglas y acrónimos	52

1. Introducción

Esta publicación presenta las características principales del proceso de codificación de las preguntas abiertas¹ relevadas en el Censo Nacional de Población, Hogares y Viviendas 2022 (Censo 2022). La tarea estuvo a cargo de la Coordinación de Clasificadores y Nomenclaturas (CCN), perteneciente a la Dirección Nacional de Metodología e Infraestructura Estadística (DNMIE) del INDEC.

En el Censo 2022, hubo tres preguntas abiertas incluidas en el cuestionario de Población:

- **La pregunta 15, opción 9, sobre el país de nacimiento.**
- **La pregunta 23, sobre el pueblo indígena u originario.**
- **La pregunta 34, sobre la rama de actividad económica.**

También se dan a conocer las metas planteadas para cada etapa de la codificación, la construcción de los instrumentos, los procedimientos metodológicos necesarios para lograrlos y los resultados de su aplicación.

La CCN no solo tuvo la responsabilidad de organizar y coordinar el proceso de codificación de las tres preguntas sino también de adaptar la Clasificación Nacional de Actividades Económicas (CNAE) y elaborar el nomenclador de pueblos indígenas (NPI) para implementarlo en el proceso de codificación de las preguntas 34 y 23. En el caso de la pregunta 15, se empleó el código de países utilizado habitualmente por el INDEC, con algunas adaptaciones.

Un aspecto significativo para el análisis del proceso de codificación se relaciona con el relevamiento del Censo 2022. La pandemia desencadenada por el virus COVID-19 intensificó el uso de datos y de internet en la vida cotidiana. La realización de los censos, a nivel mundial, requirió contemplar esta nueva dimensión digital generada por el contexto sanitario y social. En este sentido, el Censo 2022 fue la primera experiencia argentina en combinar el relevamiento realizado mediante un cuestionario autoadministrado (digital, optativo para la población) con el relevamiento tradicional (presencial, por barrido territorial) que completó la cobertura. A lo largo de la publicación se describen las especificidades que este operativo implicó para la codificación.

Igual que en los dos censos de población previos y en los operativos de encuestas a hogares, se aplicó el Sistema de Codificación Informatizada (SiCI) en el proceso de codificación

1.1 El Sistema de Codificación Informatizada (SiCI)

El Sistema de Codificación Informatizada (SiCI) se desarrolló con motivo del Censo Nacional de Población, Hogares y Viviendas 2001 y desde entonces, se amplía y se actualiza. Consiste en un sistema modular que permite tratar de manera independiente los distintos clasificadores estadísticos en el contexto de distintos operativos. De acuerdo con las características del operativo a codificar, es posible adaptar estos módulos o crear nuevos adecuados para la situación².

El SiCI se ocupa primero del acondicionamiento de la información, que consta de diversas etapas de procesamiento, como por ejemplo la corrección ortográfica y la normalización y estandarización de palabras, para que llegue preparada para la codificación. Luego, empieza el proceso de decodificación de la información, que se realiza en dos etapas, una automática y otra asistida.

El objetivo principal del SiCI es codificar de manera automática, es decir, sin intervención de personas, la mayor cantidad de casos con el menor porcentaje de error y en un tiempo reducido. A esta primera etapa se la llama **codificación automática**. Los parámetros son establecidos previamente por el equipo de la CCN, por tanto, todos los casos iguales se resuelven de la misma manera.

¹ Las preguntas abiertas son las que no están limitadas a opciones predefinidas y que permiten a los respondentes dar respuestas detalladas en texto libre.

² El desarrollo del SiCI fue impulsado y coordinado por el Sistema Nacional de Nomenclaturas, antecesor de la actual CCN. Fue un trabajo que reunió a diversos sectores del INDEC a lo largo de aproximadamente 8 años, y es un desarrollo propio de este Instituto, que se actualiza continuamente.

Existen distintos métodos de codificación automática que se aplican según las características de la información a tratar.

Los más importantes son el de autofrase, en el cual el sistema halla una frase escrita exactamente igual que en el diccionario de codificación y le asigna automáticamente el código que le corresponde, y el de microprocesos, en el cual se asigna un código con un grado razonable de certidumbre solo si se cumplen determinadas condiciones.

Para el Censo 2022, se utilizó el **método de autofrase** en las tres preguntas abiertas; en el caso de la rama de actividad, se agregaron algunos microprocesos para algunas respuestas en particular.

Los casos que no se codifican automáticamente se resuelven en una segunda fase de **codificación asistida**, donde el sistema presenta ayudas y filtros que asisten a las personas codificadoras que reciben los casos de forma automática desde una base ubicada en una red interna. Por otra parte, cada persona codificadora recibe un caso con toda la información necesaria para codificar por vez.

El SiCI agrupa las frases iguales y crea **frases únicas** para que cada una se presente una sola vez al codificador, lo cual reduce la carga de trabajo y evita el sesgo que podría suponer que una misma frase fuera codificada por distintos codificadores en diferentes instancias. Una vez establecido un código, el sistema asigna el mismo a todos los casos iguales.

La información llega ordenada a la línea de codificación. Los administradores establecen los criterios para ello, de modo que los casos aparezcan ordenados por afinidad de las variables complementarias, como los casos de una misma categoría ocupacional, y por el orden alfabético de las respuestas de las personas censadas a las preguntas abiertas. Esto permite que los casos similares se sucedan sin interrupción y que el codificador no tenga que volver a pensar más tarde una resolución que ya había encontrado.

El sistema incorpora una tabla de validación de códigos. De esta manera, si el codificador ingresa un código no válido, el sistema lo rechaza. Cabe señalar que, aunque no impide que sea adjudicado un código válido pero incorrecto, la tabla de validación reduce la cantidad de errores de codificación. En todas las etapas del SiCI, se utilizan diccionarios que son bases de datos que cumplen distintas funciones, como la corrección ortográfica, la normalización y la estandarización de palabras, la clasificación y la asignación de códigos válidos, la equivalencia con sinónimos y diferentes formas de escritura, y la codificación, entre otras.

El **diccionario de codificación** reviste particular importancia porque contiene muchos casos resueltos, es decir, frases con su código asignado. Los casos no resueltos se comparan con los del diccionario para realizar la codificación automática.

Todos los diccionarios se pueden actualizar, corregir, agregar o quitar, de acuerdo con las distintas variables u operativos. Por lo tanto, a pesar de ser un sistema determinista, el SiCI tiene una gran flexibilidad.

Cada pregunta abierta del Censo 2022 supuso un tratamiento diferente, tanto para la preparación de la codificación automática como para la organización de la segunda fase de codificación asistida. A continuación, se describe el tratamiento particular de cada una de las instancias.

2. Codificación de País de nacimiento

Se presenta el proceso de codificación de la información proveniente de la pregunta sobre el **país de nacimiento**. El bloque de preguntas involucrado en esta variable indaga el país de nacimiento de la persona censada en caso de no ser Argentina.

Las siguientes imágenes pertenecen al formulario censal de viviendas particulares del cuestionario de población:

13 ¿Nació en...	esta provincia?	☐ 1	→ Pase a 17
	otra provincia argentina?	☐ 2	
	otro país?	☐ 3	→ Pase a 15

15 ¿En qué país?	Bolivia	☐ 1	Colombia	☐ 4	Uruguay	☐ 7
	Brasil	☐ 2	Paraguay	☐ 5	Venezuela	☐ 8
	Chile	☐ 3	Perú	☐ 6	Otro país	☐ 9

Si responde **Otro país**, escriba el nombre del país.

----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

2.1 Instrumento de clasificación utilizado

Se seleccionó como base el Código de países del Instituto Nacional de Estadística y Censos (INDEC), utilizado para el registro de la migración internacional en las estadísticas de comercio exterior.³ Sin embargo, se realizaron algunas adaptaciones necesarias en virtud de las características de la información relevada en el Censo 2022.

Las adaptaciones se acordaron entre la Dirección Nacional de Estadísticas Sociales y de Población (DNESP) –en particular, la Dirección de Estadísticas Poblacionales (DEP)– y la CCN, en consulta con la Dirección de Estadísticas del Comercio Exterior (DECE). Se quitaron los códigos de las zonas francas comerciales, de los territorios vinculados a una metrópoli y de países no reconocidos por la Cancillería argentina. Además, se agregaron los antiguos códigos de países que ya no existían en 2022 y que no eran subsumibles en países actuales a esa fecha.⁴

Se incluye como anexo 1 el instrumento de clasificación final utilizado en la codificación del Censo.

2.2 Adaptación del Sistema de Codificación Informatizada

Dentro del SiCI, se creó un módulo para codificar las respuestas abiertas a la pregunta de país de nacimiento, en el que se incluyó un diccionario de codificación con los nombres de los países, las distintas formas de escritura que se detectaron en una inspección previa, y sus respectivos códigos.

Este módulo utilizó el método de autofrase, en el que el sistema asigna de manera automática el código correspondiente cuando encuentra una frase (en este caso, el nombre del país) escrita exactamente igual que alguna del diccionario de codificación.

2.2.1 Relevamiento y codificación

La recolección de la información censal fue distinta en el **Censo digital** que en el Censo por barrido territorial.

En el **Censo digital**, al llegar a esta pregunta, se le presentaba a la persona censada un listado desplegable con todos los países reconocidos en 2022. Si no encontraba su país de nacimiento, podía escribirlo.⁵

³ Disponible en <https://www.indec.gob.ar/indec/web/Nivel4-Tema-1-39-121>, columna E: "Código INDEC". O, para descarga directa, en https://www.indec.gob.ar/ftp/cuadros/territorio/codigo_paises.xls

⁴ En este Censo: Checoslovaquia, Unión Soviética y Yugoslavia.

⁵ Al realizar el análisis de la información, se encontró que las personas censadas no necesitaron utilizar esta opción porque todas pudieron ubicarse en alguna del listado.

En el **Censo por barrido territorial**, en cambio, era el censista quien anotaba lo que le dijera la persona censada. Esta modalidad tenía como opciones de respuesta predefinidas a los países limítrofes (Bolivia, Brasil, Chile, Paraguay y Uruguay), a Colombia, a Perú y a Venezuela.

Sin contar la Argentina y los ocho países recién mencionados, la cantidad de casos restantes era relativamente pequeña, por lo cual todas las respuestas a esta pregunta fueron codificadas por la CCN, sin codificadores externos.⁶

La codificación se realizó, para no demorar su inicio, sobre información sin consistir (ver el apartado 2.2.2. Revisión y cierre de las bases de datos). Por lo tanto, los números expuestos en este documento son los de este proceso de codificación, no los resultados definitivos del Censo. En la modalidad digital, se codificaron 821.896 casos, mientras que en el censo por barrido territorial, fueron 949.627.

2.2.1.1 Censo digital

En la modalidad digital de recolección de la información, se muestra en los resultados que las personas censadas no tuvieron problema al momento de responder ya que la pregunta tenía un listado desplegable de países. Por lo tanto, no hubo necesidad de codificación automática: las respuestas señaladas en el listado asignaban directamente el código del país correspondiente.

Cuadro 1. Distribución de casos de personas nacidas en otro país. Relevamiento digital, Censo 2022

	Total	Respuestas por listado	Sin respuesta
Cantidad de personas	821.896	821.185	711
Porcentaje	100,00	99,91	0,09

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

2.2.1.2 Censo por barrido territorial

En el cuestionario censal en papel, figuraban como opciones de respuesta predefinidas los países limítrofes, Colombia, Perú y Venezuela, y de esta manera, quedaron resueltos la mayoría de los casos.

Las respuestas a las preguntas abiertas fueron procesadas con el SiCI, que resultó en 578.253 casos codificados automáticamente.

Quedaron para codificación asistida solo 49.299 casos. Se codificaron, en líneas generales, sin problema salvo algunas excepciones (ver la sección 2.2.1.2.1).

Hubo 322.075 personas que respondieron haber nacido en otro país y no indicaron cuál.

Además, hubo 95.602 respuestas dobles para la misma persona censada, es decir, que se respondió tanto la pregunta cerrada como la abierta. Solo una de estas dos respuestas es parte de la base de datos final, la otra fue anulada.⁷

Cuadro 2. Distribución de casos de personas nacidas en otro país, por modalidad de codificación. Relevamiento por barrido territorial, Censo 2022

Total población codificada	Modalidad de codificación		
	Codificación automática	Codificación asistida	Sin respuesta
Cantidad de personas	578.253	49.299	322.075
Porcentaje	60,89	5,19	33,92

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

⁶ La CCN había elaborado un manual para los codificadores externos. No fue necesario utilizarlo.

⁷ La decisión recayó en la DEP. Ver el apartado 2.2.2. [Revisión y cierre de las bases de datos](#).

2.2.1.2.1 Consideraciones y criterios en la codificación asistida

Entre los 49.299 casos de codificación asistida, hubo 11.717 casos, agrupados en 1.483 frases únicas, que requirieron una atención especial. Si bien son una porción minoritaria de la codificación general, permitieron detectar algunos problemas que sirvieron en el momento del cierre de la tarea.

Los inconvenientes encontrados en este último grupo de casos fueron los siguientes:

- La pregunta debía formularse a quienes nacieron en otro país, pero algunos censistas la hicieron y consignaron a nacidos en Argentina también.⁸
- Se escribió el nombre del país a pesar de que ya figuraba en las opciones de respuesta predefinidas.
- Se consignó la ciudad o la región de nacimiento en lugar del país.
- Hubo errores de escritura en el nombre del país.
- El censista anotó el gentilicio o la etnia en lugar del país de nacimiento.

Además de estos errores asociados al registro de la información, se identificaron otras dos dimensiones de incidencia:

- Hubo personas censadas que nacieron en países que ya no existen.
- Hubo errores en el proceso de lectura y captura de la información, comunes a todas las preguntas abiertas. Siempre que fue posible, estos errores se corrigieron durante la codificación.

Las resoluciones adoptadas para solucionar estos inconvenientes fueron:

- *Si se contestó que había nacido en Argentina:*

Se lo codificó con el código de Argentina y quedó excluido del universo de la variable.

- *Si se contestó que había nacido en algún país limítrofe, Colombia, Perú o Venezuela, escribiendo el nombre del país en lugar de usar la opción de pregunta cerrada:*

Se lo codificó con el código correspondiente.

- *Si se contestó únicamente la ciudad o región de nacimiento:*

Se trató de localizar a qué país pertenecía esta ciudad o región. Cuando pertenecía inequívocamente a un país, se la codificó allí. Sin embargo, hay nombres de ciudades o regiones que se repiten en dos o más países distintos. En esos casos, se adoptaron los siguientes criterios, en orden:

- 1) Si en un país ese nombre resulta mucho más importante que en los otros, se codificó en ese primer país (por ejemplo: Cochabamba, que es una ciudad importante en Bolivia y un pequeño pueblo en Perú, se codificó como Bolivia).
- 2) Si refiere a una ciudad o región argentina y a una de otro país cuya relación resulta más indirecta, se codificó como Argentina (por ejemplo: Baradero se codificó como Argentina).
- 3) Si el nombre aparece en más de un país del mismo continente, se codificó ese continente (por ejemplo: "Itapúa", que puede referir a "Itapúa", en Paraguay, o a "Itapúa", en Brasil, se codificó como "América").
- 4) Si el nombre aparece en más de un país de distintos continentes, se anotó como no codificable (por ejemplo: Trujillo, que existe en Perú y en España).

- *Si hubo errores de escritura en el nombre del país:*

⁸ Ya sea por mencionar directamente el país o por citar alguna localidad, ciudad o provincia.

Cuando fue posible, se corrigió y se codificó en el país correspondiente. En algunas ocasiones, fue simple (por ejemplo: en “Velgica” = Bélgica), pero en otras, requirió tiempo e ingenio para descifrar cuál era el país (por ejemplo: “Chucuslabia” = Yugoslavia).⁹

- Si se contestó el gentilicio o la etnia en lugar del país:

Esto, en general, no supuso un problema (por ejemplo: “laosiano” = Laos). Pero en un caso como “árabe”, que puede remitir a varios países de Asia y África, se optó por dejarlo como no codificable.

- Si se contestó algún país que ya no existe:

Los países antiguos cuyo territorio coincide o está incluido en el territorio actual de otro se codificaron como en el territorio actual (por ejemplo: Guinea Francesa ahora es Guinea, Tanganika ahora es Tanzania).

En otros casos, siempre que no pudiera señalarse el país actual, se optó por usar los códigos vigentes en 1989, que quedaron vacantes en las clasificaciones de países actuales. Así, resultaron los siguientes casos:

408: Ex Checoslovaquia

432: Ex Yugoslavia

601: Ex Unión Soviética

2.2.2 Revisión y cierre de las bases de datos

La DNESP (particularmente, la DEP), después de haber consultado con DNMIE y con la CCN, realizó el proceso de depuración y el análisis de consistencia de las bases de datos una vez completado el proceso de codificación.

Se encontraron casos duplicados, ya que algunas personas habían contestado tanto en el Censo digital como en el Censo por barrido territorial. Además, en el Censo por barrido territorial algunas personas que contestaron la opción de pregunta cerrada, también dieron una respuesta abierta, y, a veces, estas dos respuestas no coincidían. En ambas situaciones (personas duplicadas y respuestas dobles), la decisión final acerca de cuál era la respuesta reconocida como válida recayó en la DEP.

Por estos motivos, los cuadros presentados previamente en este informe son los de la codificación, no los datos definitivos del Censo. Los resultados definitivos son los publicados con los datos oficiales del Censo 2022.¹⁰

⁹ Adicionalmente, se hace notar que la necesidad, en el cuestionario en papel, de escribir los nombres de los países no contemplados en las opciones de respuesta predefinidas, llevó a que se consignaran en algunos casos respuestas que no se correspondían con nombres de países. Por ejemplo, alguien declaró que había nacido en “Eslania”, que es un país ficticio. Estos casos fueron considerados no codificables.

¹⁰ Disponibles en https://www.indec.gob.ar/ftp/cuadros/poblacion/censo2022_migraciones.pdf

Anexo 1: Código de países

Algunas consideraciones:

- Se utilizan nombres de uso común que permiten identificar a los países, no sus nombres oficiales.
- El primer dígito de este Código de países indica el continente.
- Se usaron los códigos terminados en 98 para los casos en los cuales no se podía determinar el país concreto, pero sí el continente:

198: Indeterminados - No codificables de África

298: Indeterminados - No codificables de América

398: Indeterminados - No codificables de Asia

498: Indeterminados - No codificables de Europa

598: Indeterminados - No codificables de Oceanía

998: Indeterminados - No codificables, ni siquiera por continente

En caso de tabulados por continente, se sugirió agrupar los casos de Ex Unión Soviética con los países europeos.

CÓDIGO	PAÍS
101	Burkina Faso
102	Argelia
103	Botswana
104	Burundi
105	Camerún
107	República Centroafricana
108	Congo
109	República Democrática del Congo
110	Côte d'Ivoire
111	Chad
112	Benin
113	Egipto
115	Gabón
116	Gambia
117	Ghana
118	Guinea
119	Guinea Ecuatorial
120	Kenya
121	Lesotho
122	Liberia
123	Libia
124	Madagascar
125	Malawi
126	Malí
127	Marruecos
128	Mauricio
129	Mauritania
130	Níger
131	Nigeria
132	Zimbabwe
133	Rwanda

CÓDIGO	PAÍS
134	Senegal
135	Sierra Leona
136	Somalia
137	Swazilandia
139	Tanzania
140	Togo
141	Túnez
142	Uganda
144	Zambia
149	Angola
150	Cabo Verde
151	Mozambique
152	Seychelles
153	Djibouti
155	Comoras
156	Guinea Bissau
157	Santo Tomé y Príncipe
158	Namibia
159	Sudáfrica
160	Eritrea
161	Etiopía
162	Sudán
163	Sudán del Sur
198	Indeterminado (África)
200	Argentina
201	Barbados
202	Bolivia
203	Brasil
204	Canadá
205	Colombia
206	Costa Rica

CÓDIGO	PAÍS
207	Cuba
208	Chile
209	República Dominicana
210	Ecuador
211	El Salvador
212	Estados Unidos
213	Guatemala
214	Guyana
215	Haití
216	Honduras
217	Jamaica
218	México
219	Nicaragua
220	Panamá
221	Paraguay
222	Perú
223	Puerto Rico
224	Trinidad y Tobago
225	Uruguay
226	Venezuela
232	Suriname
233	Dominica
234	Santa Lucía
235	San Vicente y Las Granadinas
236	Belice
237	Antigua y Barbuda
238	San Cristóbal y Nevis
239	Bahamas
240	Granada
241	Antillas Holandesas
242	Aruba

CÓDIGO	PAÍS
298	Indeterminado (América)
301	Afganistán
302	Arabia Saudita
303	Bahrein
304	Myanmar
305	Bhután
306	Camboya
307	Sri Lanka
308	Corea Democrática y Popular
309	Corea
310	China
312	Filipinas
313	Taiwan
315	India
316	Indonesia
317	Iraq
318	Irán
319	Israel
320	Japón
321	Jordania
322	Qatar
323	Kuwait
324	Laos
325	Líbano
326	Malasia
327	Maldivas
328	Omán
329	Mongolia
330	Nepal
331	Emiratos Árabes Unidos
332	Pakistán
333	Singapur
334	Siria
335	Tailandia
337	Viet Nam
341	Hong Kong (región administrativa especial de China)
344	Macao (región administrativa especial de China)
345	Bangladesh
346	Brunei
348	Yemen

CÓDIGO	PAÍS
349	Armenia
350	Azerbaiyán
351	Georgia
352	Kazajstán
353	Kirguistán
354	Tayikistán
355	Turkmenistán
356	Uzbekistán
357	Palestina
358	Timor Leste
398	Indeterminado (Asia)
401	Albania
404	Andorra
405	Austria
406	Bélgica
407	Bulgaria
409	Dinamarca
410	España
411	Finlandia
412	Francia
413	Grecia
414	Hungría
415	Irlanda
416	Islandia
417	Italia
418	Liechtenstein
419	Luxemburgo
420	Malta
421	Mónaco
422	Noruega
423	Países Bajos
424	Polonia
425	Portugal
426	Reino Unido de Gran Bretaña e Irlanda del Norte
427	Rumania
428	San Marino
429	Suecia
430	Suiza
431	Vaticano
435	Chipre

CÓDIGO	PAÍS
436	Turquía
438	Alemania
439	Belarús
440	Estonia
441	Letonia
442	Lituania
443	Moldova
444	Rusia
445	Ucrania
446	Bosnia y Herzegovina
447	Croacia
448	Eslovaquia
449	Eslovenia
450	Macedonia
451	República Checa
452	Serbia y Montenegro
453	Montenegro
454	Serbia
498	Indeterminado (Europa)
501	Australia
503	Nauru
504	Nueva Zelandia
505	Vanuatu
506	Samoa
512	Fiji
513	Papua Nueva Guinea
514	Kiribati
515	Micronesia
516	Palau
517	Tuvalu
518	Islas Salomón
519	Tonga
520	Islas Marshall
521	Islas Marianas
598	Indeterminado (Oceanía)
998	Indeterminado (continente)
999	Ignorado

Códigos especiales:

CÓDIGO	PAÍS
408	Ex Checoslovaquia
432	Ex Yugoslavia
601	Ex Unión Soviética

3. Codificación de Pueblo indígena u originario

En esta sección se describe el proceso de codificación realizado para el bloque de preguntas sobre pueblos indígenas u originarios (PPII) del Censo 2022, en sus dos modalidades de relevamiento: digital y barrido territorial. De las tres preguntas sobre PPII, dos son las relevantes para la codificación:

“¿Se reconoce indígena o descendiente de pueblos indígenas u originarios?” (P22)

“¿De qué pueblo indígena u originario?” (P23)

El objetivo es desarrollar una síntesis de la metodología de trabajo que se utilizó para codificar las respuestas a estas preguntas y dejar un registro técnico de los procesos de trabajo que esté disponible para futuros relevamientos u operativos.

Se presentan las fases previas al procesamiento de la información, la elaboración del nomenclador de pueblos indígenas (NPI), las decisiones metodológicas tomadas a lo largo del trabajo, los resultados de la codificación para las dos modalidades del relevamiento censal, el control de calidad de lo codificado y, finalmente, las conclusiones generales.

3.1 Fases previas a la codificación

Con la incorporación de la temática indígena en los relevamientos estadísticos del INDEC a partir del Censo 2001, se pudo establecer un listado de pueblos y su distribución territorial. Para el Censo 2022, los objetivos de la CCN y de la DNESP abarcaron no solo la publicación de los cuadros con los resultados censales, sino también la elaboración de un nomenclador propio que dejara registro del instrumento estadístico con el que se procesó la información y sirviera como instrumento base a partir del cual desarrollar nuevas versiones para los relevamientos venideros.¹¹

3.1.1 La elaboración del nomenclador de pueblos indígenas (NPI)

La elaboración del nomenclador implicó revisar fuentes propias del INDEC¹² y de otras organizaciones y organismos¹³ y convocar a las autoridades del Instituto Nacional de Asuntos Indígenas (INAI), organismo que promueve las políticas públicas para la población indígena en nuestro país que, en este caso, aportó el conocimiento necesario para su validación.

Al comienzo de las tareas en 2019, el INAI remitió al INDEC un listado de 39 pueblos con personería jurídica, es decir, reconocidos formalmente. Sin embargo, tras analizar los resultados de la Segunda Prueba Piloto (II PP) del Censo 2022, se concluyó que el NPI debía ampliar las posibilidades de registro dada la aparición de respuestas que mencionaban comunidades, sinónimos u otras formas de denominación o escritura. Estas respuestas debían ser tomadas en cuenta respetando el proceso de reconocimiento o visibilidad pública que tuviera o pudiera tener el pueblo indígena (PI) referido.

De esta forma, se reagruparon los pueblos teniendo en cuenta el proceso abierto y dinámico de visibilización y revalorización indígena, incorporando las especificidades¹⁴ de cada pueblo donde correspondiera y contemplando las referencias geográficas, entre otros aspectos. Esto permitió incorporar nuevos nombres al nomenclador y relacionar comunidades, aldeas, apellidos y parajes con pueblos de forma inequívoca. La incorporación de estas descripciones

¹¹ “Es relevante señalar brevemente el fenómeno de ‘reemergencia indígena’ o ‘etnogénesis’, que refiere a un proceso dinámico de revalorización de la identidad indígena, cuyos resultados (como ser la ‘visibilización’ de un pueblo o etnia previamente invisibilizado) pueden variar de un período censal a otro. Por lo tanto, no se deben tomar como definitivas las categorías de los pueblos indígenas, sino que es necesario la actualización constante de las mismas” (INDEC, s.f.).

¹² Censos Nacionales de Población, Hogares y Viviendas de los años 2001 y 2010; Encuesta Complementaria de Pueblos Indígenas (ECPI) 2004-2005; y Segunda Prueba Piloto del Censo Nacional de Población, Hogares y Viviendas 2022.

¹³ El Registro Nacional de Comunidades Indígenas del INAI, el Mapa de Pueblos Naciones Originarias en la actualidad del Encuentro Nacional de Organizaciones Territoriales de Pueblos Originarios (ENOTPO) y el listado de Pueblos del Equipo Nacional de Pastoral Aborigen (ENDEPA).

¹⁴ En los distintos relevamientos, se detectaron menciones de especificidades del interior de algunos PI. Estas especificidades suelen originarse en la reivindicación de las identidades locales, por lo que se decidió conservarlas bajo el tercer dígito del nomenclador. Por ejemplo, se puede decir que el Pueblo Nación Diaguita está conformado por los pueblos Diaguita Amaicha, Diaguita Calchaquí, Diaguita Quechua, Diaguita Ingamana, Diaguita Quilmes, Diaguita Tolombón y Diaguita Colastiné.

permitió mejorar el análisis de las respuestas y, como consecuencia, su codificación. Así y como se observará más adelante, el NPI se fue completando en las distintas etapas del proceso de trabajo.

3.1.2 Decisiones metodológicas

La población objetivo se conformó con las personas que se reconocieron como indígenas o descendientes de pueblos indígenas latinoamericanos que habitaron el territorio antes de la conquista y colonización, previa conformación de los estados nacionales. Se configuró, en primer lugar, con base en todas las respuestas afirmativas que contenía la P22. En segundo lugar, se tomaron en cuenta todas las respuestas en la P23, salvo que refirieran a países, lugares o PPII por fuera de América Latina. Esas respuestas no fueron consideradas como parte del universo de la pregunta sobre pueblos originarios.

En la II PP se observó que algunas descripciones contenían información insuficiente para asignar un código de PI, por ejemplo, “pueblo aborigen”, “comunidad”, “desconocido”. Tales descripciones no permitían identificar a un pueblo indígena de forma inequívoca. Además, se registraron respuestas relativas a PPII de Latinoamérica no incluidos en el nomenclador. Para estos casos, se empleó el código 970.

A las respuestas relativas a países o PPII que no fueran latinoamericanos, se las codificó con el código 980. Como ejemplos, se pueden mencionar “gitano”, “Italia”, “sioux”. También fueron codificadas con este código las menciones relativas al reconocimiento afrodescendiente¹⁵. Todos estos casos fueron excluidos de la población objetivo de las P22 y P23¹⁶.

Se incorporó el código provisorio 950 con el fin de marcar descripciones de posibles sinónimos o especificidades que podían remitir a PPII, o bien, PPII que no habían sido definidos anteriormente, que fueron resueltos por el equipo INAI-INDEC en una segunda instancia (ver el anexo 2, “Pautas de codificación”).

Por último, para los casos sin respuesta (en blanco), se utilizó el código 990.

3.2 El relevamiento: la captación de pueblos indígenas en las dos modalidades del censo

El Censo 2022 tuvo dos modalidades de relevamiento: **digital** y por **barrido territorial**. El digital se relevó entre el 16 de marzo y el 18 de mayo de 2022; el barrido territorial, el 18 de mayo de 2022.

Si bien la formulación tanto de la P22 como de la P23 fue la misma en ambas modalidades, el relevamiento digital se respondía con un cuestionario autoadministrado por internet. La población debía colocar una marca en una de las opciones predefinidas y propuestas en un listado desplegable. Si no se identificaba un PI, podía marcar la opción “Otros” y completarla manualmente.

Para el operativo del Censo en su modalidad por barrido territorial, la captación fue a la manera tradicional: quienes censaban debían registrar la totalidad de las respuestas por escrito en el espacio de casilleros correspondiente, tal como se observa en la imagen que sigue:

22 ¿Se reconoce indígena o descendiente de pueblos indígenas u originarios?

Sí	☐ 1
No	☐ 2
Ignorado	☐ 9

→ Pase a **25**

23 ¿De qué pueblo indígena u originario?

----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

Ignorado ☐ 9

¹⁵ El equipo de la DNESP sumó estos casos al universo de la P25, que indaga explícitamente por este grupo poblacional.

¹⁶ En total se encontraron 9.354 casos (0.65%) con este código.

Se inició el trabajo de codificación sin el proceso de validación y consistencia¹⁷, con el objetivo de acortar los tiempos para disponer de la información final para su publicación. Por lo tanto, los resultados de la codificación presentados en este informe no deben confundirse con los resultados definitivos publicados.

3.3 La codificación de la pregunta de Pueblos indígenas

El proceso de codificación implicó una diferencia según la modalidad en la que se obtuvieron las respuestas. El relevamiento digital contó con la información digitalizada desde el inicio, sin que fuera necesaria una lectura externa. En cambio, la modalidad por barrido territorial necesitó de un proceso de escaneo de las respuestas para la constitución de la base final.

Para procesar la información se utilizó el SiCI, que permite codificar las declaraciones censales de forma automática. Los casos que no pudieron ser resueltos de esta manera se codificaron de forma asistida. En esta instancia, el trabajo se realizó a través del análisis interpretativo de un equipo de codificación, con pautas metodológicas constituidas a tal efecto y asistencia técnica durante el proceso de codificación.

Las respuestas de la modalidad digital estuvieron disponibles terminado el relevamiento, lo que posibilitó que la CCN lo codificara durante junio de 2022. Los casos no codificados de forma automática se resolvieron de manera asistida, como se verá más adelante.

Las respuestas obtenidas mediante el barrido territorial estuvieron disponibles para codificar en abril de 2023, luego de la primera lectura de la información a través de escaneo, fue necesaria una revisión y relectura posterior de los cuestionarios de algunas provincias ya que se detectaron errores. Esto explica, en gran parte, el rezago de la codificación por barrido con respecto a la modalidad digital.

A continuación, se desarrollará el proceso de codificación en cada una de sus modalidades.

3.3.1 Codificación del relevamiento digital

En el procesamiento del relevamiento digital, 411.310 personas se reconocieron pertenecientes a PPII.

Cuadro 3. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, según modalidad de codificación. Relevamiento digital, Censo 2022

Modalidad de codificación	Población que se reconoce indígena o descendiente de pueblos indígenas u originarios	Porcentaje
	Total	%
Total	411.310	100,00
Automática	382.474	92,99
Asistida	28.836	7,01

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

Tal como podía esperarse, la mayor parte de la codificación del relevamiento digital (poco más de 9 de cada 10 respuestas) se hizo de forma informatizada debido, en gran medida, al listado de opciones predefinidas del desplegable. Además, el SiCI también codificó algunas de las respuestas abiertas de la opción “Otros” del cuestionario. Luego, la codificación asistida se abocó a las respuestas de esa misma opción que no pudieron ser resueltas de manera automática.

En la preparación de los casos para la codificación asistida, se utilizó un procedimiento informático llamado “frase única” que reduce la cantidad de casos a codificar. El procedimiento consiste en identificar los casos y las frases iguales o repetidos en el conjunto. Cuando un codificador resuelve uno de esos casos reiterados, el sistema le asigna

¹⁷ La validación y consistencia, posterior a la codificación a cargo de la DNESP, podía modificar el tamaño de la población objetivo.

automáticamente el mismo código a los demás. Gracias a este método, se pudo reducir el 20% de los casos reales que hubo que codificar de manera asistida.

La CCN consideró que la codificación de ese porcentaje de frases únicas se podría resolver sin necesidad de recurrir a un equipo de codificación externo, tal como se había estimado en el inicio. Se llevó adelante esta tarea a lo largo de junio de 2022. Esto se decidió y realizó con el aval del INAI, al que también se lo consultó ante casos dudosos pendientes de resolución.

La codificación asistida sirvió para seguir completando el diccionario del SiCI, agregando casos nuevos –es decir, pueblos indígenas no registrados previamente en el nomenclador– validados y actualizando el nomenclador. Esto se detallará en el apartado 3.4.

El detalle de los resultados de la codificación es el siguiente:

Cuadro 4. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, por código asignado, según modalidad de codificación. Relevamiento digital, Censo 2022

Código asignado	Codificación automática		Codificación asistida	
	Total	%	Total	%
Total	382.474	100,00	28.836	100,00
010 a 580: pueblo indígena ⁽¹⁾	337.250	88,18	17.082	59,24
970: información insuficiente	44.161	11,55	11.716	40,63
990: en blanco/sin respuesta	1.063	0,27	38	0,13

⁽¹⁾ En esta categoría se agruparon todas las respuestas que indican la pertenencia a un PI determinado. Ver en Anexo el Nomenclador de Pueblos Indígenas.

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

Gracias a las opciones predefinidas, la codificación automática pudo asignarle un código de PI a la mayoría de las respuestas (88,18%). También resolvió con un código las frases que se habían caracterizado como “información insuficiente” y “sin respuesta”.

En el caso de la codificación asistida, casi el 60% recibió un código de PI. Se trató como información insuficiente para codificar al 40% restante, y se le asignó el código 970. Esto significa que los textos respondidos en la opción “Otros” no permitieron la asignación inequívoca de un código del NPI.

Los casos sin respuesta en esta modalidad fueron cercanos a cero.

3.3.1.1 Control de calidad del relevamiento digital

El control de calidad de lo codificado en esta modalidad se facilitó gracias al listado desplegable de la P23. Quienes respondieron el censo de esta forma, marcaban el PI correspondiente entre todas las opciones predeterminadas. Si no lo encontraban, podían escribirlo en el espacio previsto luego de marcar “Otros” en el desplegable.

La codificación automática se hizo a través del SiCI, que leía los textos y asignaba los códigos. Al momento de revisar el resultado, se encontraron inconsistencias en la distribución de la representación de los distintos PI. Estas inconsistencias fueron subsanadas conjuntamente entre la CNN y el equipo informático.

Distinto fue el tratamiento de la codificación asistida. Como ya se indicó en el punto anterior, la CCN asumió la codificación de esta modalidad en consulta permanente con el INAI. Las dudas que surgieron en esta instancia se resolvieron más tarde, con los aportes del equipo que codificó la parte asistida de la modalidad por barrido. Efectivamente, terminado ese proceso, se consistió toda la codificación asistida del relevamiento digital con las decisiones de lo codificado por barrido. Ese procedimiento aclaró las dudas y permitió cotejar que declaraciones similares tuvieran asignado el mismo código. Así se garantizó un control sistemático del error.

3.3.2 Codificación del relevamiento por barrido territorial

A continuación, se describen las condiciones de origen de los datos a codificar y la lectura de las respuestas escaneadas. Además, se detalla el procesamiento de los datos de forma automática y asistida, así como los obstáculos que se encontraron al momento de realizar el control de calidad.

3.3.2.1 Condiciones de origen de los datos a codificar. Lectura de las respuestas

Como ya se mencionó en el apartado 3.3, una lectora escaneó las respuestas que las personas censistas registraron manualmente en el cuestionario del relevamiento por barrido. Luego, se configuró un archivo digital para su codificación. Se alojaban los archivos ordenados por provincia a los servidores utilizados para el censo para que la CCN revisara y controlara la calidad de los textos.

Gracias a este procedimiento, se detectaron inconvenientes en la legibilidad de los textos de algunas provincias. A modo de ejemplo, la tipología de los errores encontrados se puede ver en el Esquema 1.

Esquema 1. Tipología de errores encontrados en la lectura del escaneo de las respuestas a la pregunta 23 del cuestionario censal

Caracteres especiales ilegibles	Conjuntos de letras sin ningún tipo de posibilidad de lectura	Conjuntos de letras que permiten algún tipo de legibilidad e identificar una palabra codificable
	AAY JHE	NIANPYCKIELH C
= * *	BJEEWITRAIS U I	OIIISINDRIC
= * * *	CBSDAZA	PHZCTIIPRIJ
= * * * * * N	DCINVII	QIHZQVGISILM
= * * * I * L * LR *	EF IIIDAIIRE	R R RR IIII
= * * * * IIII =	FDILLAIIRIÑ	SGWTZYJCRUGI
= * * I * I	G IIW I	TINKNSR RCIII
= * * I * * R * S * *	HINGHTMAPUCHEÑGU	UCICAYFICPS
= J I I = K = * = I	I IHI	VIRBIIIZIJI
= * H I I I H = I T	JIGIIAIRINITNRJL	WUIIIIIQIWRWIII
= A = U = H * E * MLK *	KIIII ÑIIII	XUABUPEÑRWOEAEKO
= RU = * = I =	L R Z I	YBIII IYI
	MAIJWHIHE I R	ZEL PFDZG MGWNJZT

Fuente: Extracto elaborado por CCN de las bases enviadas por la empresa lectora.

Tras verificar estos problemas, se devolvió la información para que de la empresa responsable del escaneo y de la lectura la revisara y corrigiera. Si bien mejoró la legibilidad de una porción de los casos, no alcanzó a la totalidad de las respuestas.

Una vez recibidos estos archivos, la CCN los volvió a procesar con el SiCI en abril de 2023. La revisión del segundo procesamiento detectó respuestas incongruentes en cuanto al PPII y la localización. En Chubut y Neuquén, muchas personas respondieron “altiplano” y muy pocas respuestas de pueblos históricos de esas provincias. Dado que las dos provincias patagónicas son parte del territorio de esos pueblos, se decidió cotejar los archivos con las imágenes de los cuestionarios. Se verificó de manera sistemática que las respuestas de los cuestionarios eran “Mapuche”, es decir que, por problemas en la lectura, el sistema leía “altiplano” donde decía “Mapuche”. Este mismo problema se

registró en la provincia de Salta, aunque en menor proporción. Se advirtió el error, se corrigió y se extendió a todas las provincias.

Tales inconvenientes hicieron que fuera necesario volver a procesar las respuestas corregidas por segunda vez. Como consecuencia, se demoró el inicio de la codificación asistida de estas respuestas y de todas las tareas posteriores para controlar la calidad de lo codificado y consolidar la información final.

A lo anterior, también se suman algunas dificultades propias del relevamiento censal. Hubo respuestas a la P23 del tipo “Mar del Plata”, “Balcarce”, “Resistencia”, “Trelew”, etc. que surgieron como resultado, probablemente, a que la pregunta se formuló “¿De qué pueblo?” y no se explicitó “indígena u originario”, por lo que la persona censada respondía remitiendo a su localidad de origen o al lugar en el que vivía en ese momento. La CCN advirtió este equívoco en la Segunda Prueba Piloto del Censo¹⁸.

La CCN procedió a la codificación por barrido con el mismo método utilizado para la modalidad digital: en primera instancia, el procesamiento automático mediante el SiCI, cuyo diccionario estaba conformado por los nombres de los PI ya registrados en relevamientos anteriores, sinónimos, variantes validadas y los nuevos¹⁹ pueblos identificados en la modalidad digital. En segunda instancia, la codificación asistida.

El resultado, según la modalidad de procesamiento, se observa en el siguiente cuadro:

Cuadro 5. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, según modalidad de codificación. Relevamiento por barrido territorial, Censo 2022

Modalidad de codificación	Población que se reconoce indígena o descendiente de pueblos indígenas u originarios	Porcentaje
	Total	%
Total	1.007.070	100,00
Automática	726.621	72,15
Asistida	280.449	27,85

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

Se procesaron 1.007.070 respuestas. El SiCI logró asignar un código de manera automática a las tres cuartas partes, y el cuarto restante quedó para ser codificado de manera asistida por las personas codificadoras.

¹⁸ El cuestionario de la II PP de setiembre 2019 tenía la P23 pregunta formulada de esa manera. El informe de codificación de la prueba, realizado por la CCN, recomendó agregar “indígena u originario” a la formulación “¿de qué pueblo?” tras constatar las respuestas equívocas suscitadas en la observación en campo. Es probable que las respuestas en el censo fueran el resultado de la formulación acortada de la pregunta. Por otra parte, durante la codificación del relevamiento por barrido territorial, referentes del INAI recordaron notar cierta vergüenza o prejuicios en quienes censaban para realizar la pregunta, formulándola de forma incompleta, lo que también podría haber originado respuestas incorrectas.

¹⁹ Ver nota 19.

3.3.2.2 Codificación del relevamiento por barrido territorial en su modalidad automática

Cuadro 6. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, por código asignado, según modalidad de codificación automática. Relevamiento por barrido territorial, Censo 2022

Código asignado	Total	%
Total	726.621	100,00
010 a 580: pueblo indígena	327.996	45,14
970: información insuficiente	33.288	4,58
990: en blanco/sin respuesta	365.337	50,28

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

El SiCI pudo asignarle un código de PI a poco menos de la mitad de lo codificado. A la otra mitad le asignó el código previsto para los casos sin respuesta y, en menor medida, el código de información insuficiente. Si bien el bloque de preguntas admite la posibilidad de que una persona conteste reconociéndose indígena o descendiente de indígena y no responda de qué pueblo, porque puede no saberlo con certeza, la magnitud de las respuestas en blanco llevó a establecer la necesidad de un análisis específico de este subuniverso que se detalla en el apartado 3.3.2.4.

3.3.2.3 Codificación del relevamiento por barrido territorial en su modalidad asistida

Previendo que la cantidad de casos a codificar de manera asistida sería mayor que la de la modalidad digital, la CCN dispuso la organización de un dispositivo de codificación que permitiera terminar la codificación en los tiempos previstos. Dadas las características de la temática a codificar, el INDEC le solicitó nuevamente la colaboración al INAI, que eligió un grupo de referentes indígenas para llevar a cabo el trabajo.

El 18 de octubre de 2023, el equipo de la CCN, el grupo de referentes indígenas propuesto por el INAI y sus autoridades se reunieron en el INDEC con el objetivo de organizar y establecer, en conjunto, las pautas de trabajo con las que se abordaría la tarea. El equipo conformado contó con representantes de diferentes PPII: Diaguita, Mapuche, Guaraní, Huarpe, Kolla, Günün a Küna, Charrúa, Comechingón, Lule y Fiscara, quienes conocen la realidad territorial de los pueblos y comunidades de todo el país. Se acordó que el equipo trabajara ad honorem desde sus respectivas provincias y sea coordinado por la CCN.

Las tareas se iniciaron unas semanas después con un breve curso de capacitación virtual dictado por la CCN, en el que se presentaron los materiales de trabajo que se usarían en la codificación:

- **un manual con las distintas definiciones y pautas metodológicas;**
- **el NPI;**
- **ejemplos para ejercitación.**

Debido a algunos inconvenientes de agenda del equipo de referentes, el curso se dictó en tres oportunidades hasta completar la capacitación. El 100% de la información para codificar de forma asistida se redujo a un 35% de frases únicas. Ésta se fragmentó en diez lotes, respetando una propuesta de regionalización²⁰ solicitada por el equipo. Esos lotes fueron enviados el 22 de noviembre de 2023 a las y los respectivos responsables, en conformidad a lo establecido por la Ley 17.622 que regula el secreto estadístico.

²⁰ A efectos de la codificación, se constituyeron cinco regiones: NOA (Jujuy, Salta, Tucumán, Santiago del Estero, La Rioja y Catamarca), NEA (Misiones, Corrientes, Chaco, Formosa, Entre Ríos y Santa Fe), Centro (San Juan, San Luis, Mendoza, La Pampa y Córdoba), Sur (Río Negro, Neuquén, Chubut, Santa Cruz y Tierra del Fuego) y provincia de Buenos Aires y CABA. Se envió sólo la información necesaria para codificar sin ninguna referencia que pudiera violar el secreto estadístico.

El trabajo se completó en poco más de un mes²¹, con reuniones semanales para compartir dudas y elaborar criterios de resolución comunes. Como resultado más importante, se debe mencionar la aparición de respuestas relativas a pueblos que no estaban en el nomenclador, varios de los cuales se creían extintos (por ejemplo, los pueblos Wayteka y Jujuí, entre otros). A raíz de esto último, el nomenclador pasó de tener 50 PI principales al inicio del trabajo, a registrar 58 al final de la codificación asistida. Por otra parte, gracias al conocimiento de las personas referentes, se pudieron vincular varias comunidades con un pueblo indígena de forma inequívoca. A su vez, la valiosa glosa de información que el equipo registró sobre los criterios y las decisiones de codificación y su justificación serviría como material imprescindible para elaborar, posteriormente, las notas explicativas del nomenclador.

Este proceso de codificación permitió la actualización del NPI no solo con pueblos no identificados hasta el momento, sino también con el registro de otras comunidades, especificidades o sinónimos de PI no contemplados en el nomenclador utilizado en el inicio del proceso.

El resultado de la codificación de la población que se reconoce indígena o descendiente de pueblos indígenas u originarios por código asignado en su modalidad asistida es el que se muestra en el siguiente cuadro.

Cuadro 7. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, por código asignado, según modalidad de codificación asistida. Relevamiento por barrido territorial, Censo 2022

Código asignado	Total	%
Total	280.449	100,00
010 a 580: pueblo indígena	202.926	72,36
970: información insuficiente	68.131	24,29
990: en blanco/sin respuesta	9.392	3,35

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

El porcentaje de casos sin respuesta que se observa en esta modalidad corresponde a la lectura de alguna marca en la información recibida que no se verificó en la revisión de las imágenes de los cuestionarios. Por esa razón, no pudo ser resuelto por el SiCI.

Si se compara con la modalidad automática, se observa que el conocimiento de la temática por parte de las personas convocadas por el INAI, ayudó a asignar e identificar una mayor proporción de pueblos indígenas en esta modalidad.

3.3.2.4 Control de calidad del relevamiento por barrido territorial

En el caso del control de calidad del relevamiento por barrido territorial, no se dispuso de un sistema de control por pareo mediante recodificación independiente a doble ciego, como sí se hizo para la codificación de la pregunta 34, que se verá más adelante.

La revisión y control de lo codificado de forma automática no presentó mayores inconvenientes dado su funcionamiento determinístico de búsqueda de frases o palabras ya previstas en su diccionario. Además, el equipo de la CCN resolvió la porción de casos sin respuesta y corrigió los pocos errores encontrados en su totalidad.

Con respecto a la modalidad asistida, se decidió que el equipo revisara la totalidad de los casos codificados y verificara que se hubieran cumplido las pautas establecidas en el manual de capacitación presentadas en el curso.

La tarea se concentró en el control de la consistencia interna de la información codificada, procurando que los códigos aplicados para una descripción de la P23 que se repite fueran los mismos en la totalidad de los casos.

²¹ Los tiempos de codificación por parte del equipo del INAI fueron satisfactorios y coincidieron con los estimados por la CCN.

Para las respuestas en las que no se pudo identificar algún pueblo, el equipo de la CCN realizó búsquedas en la web cuando del formulario se podía inferir la existencia de un PI y luego, en caso de que así fuera, se realizaba la consulta final al INAI²².

Una vez concluida la codificación asistida, la DNESP y el INAI suministraron listados con nombres de comunidades indígenas que sirvieron para mejorar la calidad de la codificación, en especial para los casos que hasta ese momento tenían el código de información insuficiente (970).

Teniendo en cuenta la importante magnitud de no respuesta, la DNESP decidió revisar esos casos contrastándolos con las imágenes de los cuestionarios para recuperar respuestas efectivas y que el escaneo posiblemente no los haya leído.

Un equipo del INDEC conformado ad hoc revisó un total de 45.000 imágenes de los cuestionarios e ingresó manualmente los textos encontrados. De éstos, más de la mitad tenían respuesta. Con este procedimiento fue posible reducir en un 7% los casos en blanco.

3.4 Resultados del proceso de codificación

En el proceso de codificación, se reconocieron pertenecientes a pueblos indígenas 1.418.380 personas. El 29% de ellas respondió el cuestionario de forma digital y el 71% lo hizo mediante el operativo del barrido territorial.

Cuadro 8. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, según modalidad de relevamiento censal. Censo 2022

Modalidad de relevamiento	Población que se reconoce indígena o descendiente de pueblos indígenas u originarios	Porcentaje
	Total	%
Total	1.418.380	100,00
Digital	411.310	29,00
Por barrido territorial	1.007.070	71,00

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

Mientras que alrededor del 50% de la población total respondió el censo en su modalidad digital, en el caso de la población indígena, fue el 29%. Teniendo en cuenta que esta modalidad de relevamiento del censo fue la primera experiencia a nivel nacional, el balance de los resultados obtenidos es positivo. Dicho balance se apoya en la proporción de respuestas válidas registradas para esta variable, que fue mayor en la modalidad digital que en el barrido territorial.

²² También se consultó al INAI en los casos que aparecían términos como "Comunidad" y "Colonia" que no contaban con un código de un PI por tratarse, sobre todo en el caso de las comunidades, de una organización de vida frecuente de los pueblos que utilizan para identificarse.

Cuadro 9. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, según código asignado. Censo 2022

Código asignado	Población que se reconoce indígena o descendiente de pueblos indígenas u originarios	Porcentaje
	Total	%
Total	1.418.380	100,00
010 a 580: pueblo indígena	885.254	62,41
970: información insuficiente	157.296	11,09
990: en blanco/sin respuesta	375.830	26,50

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

En lo que refiere al contenido de las respuestas, a casi dos tercios se le pudo asignar un código de pueblo indígena específico según el nomenclador. El resto se repartió entre “información insuficiente para codificar” (descripciones imprecisas que en su mayoría refieren a lugares o que sencillamente no permiten identificar un pueblo indígena) y la no respuesta, que refiere a reconocerse indígena o descendiente en la P22 y dejar vacía la P23.

Aquí, es necesario aclarar que la falta de respuesta en la P23 no se explica solo por problemas de relevamiento o capacitación de quienes censaron o el escaneo de los cuestionarios. Una porción importante de la población que respondió que se reconoce indígena o descendiente no tiene conocimiento certero de a qué pueblo pertenece. En algunos casos, lo expresaron explícitamente: “no sé de qué pueblo”, “desconozco”, etc. Con esto, se quiere resaltar la importancia de la respuesta positiva en P22, acorde con el avance en el trabajo de revalorización de la identidad indígena que llevan adelante las distintas personas representantes de los PI.

La constatación empírica de este resultado puede observarse en el cuadro que sigue, en el que se comparan los resultados totales entre quienes contestaron reconociéndose como indígenas o descendientes en Censo 2022²³ y los que lo hicieron en los dos relevamientos estadísticos anteriores que también lo preguntaron²⁴. Cabe aclarar que tanto la ECPI 2005 como el Censo 2010 captaron a esta población a través de una muestra.

Cuadro 10. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, según ECPI 2005, Censo 2010 y Censo 2022

	ECPI 2005	Censo 2010	Censo 2022 ⁽¹⁾
Población que se reconoce indígena o descendiente de pueblos indígenas u originarios	600.329	955.059	1.418.380

⁽¹⁾ Datos sin consistir.

Fuente: INDEC, CCN. Elaboración propia con base en ECPI 2005, Censo 2010 y Censo 2022.

En 2022, se reconocen pertenecientes o descendientes de PPII 136,3% más personas que en 2005 y un 48,5% más que en 2010.

Por último, se destaca la posibilidad de ampliar el nomenclador durante el proceso de codificación de cada una de las modalidades del censo.

²³ El número para el Censo 2022 es preliminar.

²⁴ Encuesta Complementaria de Pueblos Indígenas 2005 y el Censo Nacional de Población, Hogares y Viviendas 2010.

Los pueblos y especificidades que surgieron durante la codificación del relevamiento digital fueron los siguientes:

- Se incorpora el pueblo Alakaluf al nomenclador.
- El pueblo Diaguita Cacano deja de estar en el grupo Diaguita y pasa a tener su propio código.
- Se incorpora el pueblo Diaguita Ingamana como especificidad dentro del código Diaguita.
- Se incorporan especificidades dentro del pueblo Guaraní y se reagrupan bajo este código el Ava Guaraní y el Tupí Guaraní.
- Las anteriores especificidades Araucano y Nehuenche del grupo Mapuche pasan a ser sinónimos.
- Se incorpora el pueblo Querandí al nomenclador.
- El pueblo Mataco deja de ser una especificidad dentro del pueblo Wichi por tratarse de una descripción peyorativa y se conserva en el diccionario de codificación para no perder la información.

Las surgidas durante el relevamiento por barrido territorial fueron:

- Se incorporaron al nomenclador ocho PPII: Wayteca, Haush/Maneken, Mak'a, Minuán, Ansilta, Churumata, Jujuí, Michilingüe.
- Caraca y Kaingang se agregan como sinónimos del pueblo Chana.
- Se suma el Pueblo Colastiné como especificidad del pueblo Diaguita.
- Guácara y Santa Ana de las Guácaras se incorporan como sinónimos del pueblo Guaycurú.
- Se suman sinónimos al pueblo Günün a Küna: Cangapol, Diuhet, Gastre Tehuelche, Chechehet, Taluhet, Pampas, Pampa Araucanizado, Tehuelche del norte (de Patagonia norte).
- Borogas y Catriel Pampas son sinónimos del pueblo Mapuche.
- Masacre Rincón Bomba, Achalai y Tres Isletas se agregan como sinónimos del pueblo Pilagá.
- Se agrega el pueblo Arancuelche como otra forma de referirse al pueblo Ranquel.

El conocimiento del grupo de referentes que propuso el INAI dio fundamento a estas modificaciones. Sin su conocimiento experto en la materia, el nivel de detalle de la codificación no hubiera sido posible.

3.5 Conclusiones

Desde el 2001, el relevamiento de la pertenencia o descendencia de un pueblo indígena u originario ha dado cuenta de la importancia de su medición en el operativo censal. El incremento registrado entre censos de quienes así se reconocen, como puede verse en el apartado 3.4, ratifica esa afirmación (cuadro 10).

El Censo 2022 implementó un operativo con dos modalidades de relevamiento: el barrido territorial y la modalidad digital. La modalidad por barrido mostró una mayor proporción de respuestas de quienes se reconocen parte de los PI (71%) con relación a la modalidad digital (29%), tal como se mostró en el cuadro 6.

Si bien en el relevamiento por barrido se obtuvieron más respuestas, un porcentaje alto no pudieron ser codificadas, se encontraban vacías o en blanco. Una parte de esto podría explicarse por un aumento en la autoidentificación indígena sin conocer con certeza a qué pueblo específico.

Por su parte, el uso de un desplegable con nombres de pueblos predeterminados en el relevamiento digital y la posibilidad de contar con un espacio para escribir en caso de ser de un pueblo ausente en esa lista tuvo mejores resultados, con bajo porcentaje de no respuesta y de información insuficiente para codificar. Esto se observa en el siguiente cuadro.

Cuadro 11. Población que se reconoce indígena o descendiente de pueblos indígenas u originarios, por código asignado, según modalidad de relevamiento censal. Censo 2022

Código asignado	Modalidad digital		Modalidad por barrido territorial	
	Total	%	Total	%
Total	411.310	100,00	1.007.070	100,00
010 a 580: pueblo indígena	354.332	86,15	530.922	52,72
970: información insuficiente	55.877	13,59	101.419	10,07
990: en blanco/sin respuesta	1.101	0,27	374.729	37,21

Fuente: INDEC, base preliminar del Censo 2022. Coordinación de Clasificadores y Nomenclaturas. Dirección Nacional de Metodología e Infraestructura Estadística.

De estos hechos podemos concluir que, en la medida que mejore la conectividad de la población y las campañas previas de capacitación, la opción de autorrespuesta será muy efectiva en cuanto a la calidad de lo captado.

Sin embargo, mientras la captación por barrido territorial con censistas continúe siendo la forma determinante de llegar a la totalidad de la población que no pueda contestar por otros medios, deberán mejorarse los siguientes aspectos:

- La captación y calidad de las respuestas en todos sus aspectos. Para ello, es necesario insistir en la participación sistemática de referentes indígenas en las diferentes etapas del relevamiento censal. En las situaciones donde esto no sea posible, es importante que cada censista cuente con una buena capacitación y sensibilización sobre la temática, además de recibir ejercitación práctica sobre estas preguntas en el cuestionario.
- La lectura de las respuestas abiertas. Tal como se observó en el apartado 3.3.2.1 se debe mejorar la lectura y escaneo realizados

Para finalizar, la participación del INAI fue de suma importancia. El INDEC aseguró el cumplimiento de la normativa constitucional e internacional²⁵ sobre la integración de los PI en todas las cuestiones que los involucren. Tal como se menciona en el punto 3.3.2.3, esto dio por resultado la identificación de numerosos pueblos, incluyendo algunos que se creían extintos, lo que permitió que, a la fecha, se disponga de un nomenclador con 58 PI ordenados y agrupados con sus especificidades.

La CCN tiene por delante el desafío de elaborar un nuevo clasificador de PI que incorpore las dimensiones analíticas aprendidas a lo largo del trabajo descrito en esta publicación.

²⁵ La Constitución de la Nación Argentina, artículo 75, inciso 17; la Ley 23.302 de creación del INAI; y el Convenio 169 de la Organización Internacional del Trabajo sobre Pueblos Indígenas y Tribales en Países Independientes (Ley 24.071); entre otras.

Anexo 2. Nomenclador de pueblos indígenas

Código	Pueblo Indígena	Código	Pueblo Indígena	Código	Pueblo Indígena
010	Atacama	172	Ava Guaraní	350	Qom/Toba
020	Alakaluf	173	Tupí Guaraní	360	Quechua
030	Aoniken	180	Guarayo	370	Querandí
040	Avipón	190	Guaycurú	380	Ranquel
050	Aymara	200	Günün A Küna	390	Sanavirón
060	Chaná	210	Huarpe	400	Selk'nam/Ona
070	Chané	220	logys	410	Tapiete
080	Charrúa	230	Isoceño	420	Tastil
090	Chicha	Kolla		430	Tehuelche
100	Chorote	241	Kolla	440	Tilian
110	Chulupí/Nivacle	242	Kolla Diaguita	450	Toara
120	Comechingón	243	Kolla Quechua	460	Tonokoté
130	Corundí	250	Kolla Atacameño	470	Vilela
Diaguita		260	Lule	480	Weenhayek
141	Diaguita	270	Lule Vilela	490	Wichi
142	Diaguita Amaicha	Mapuche		500	Yagán
143	Diaguita Calchaquí	281	Mapuche	510	Wayteca/Chono
144	Diaguita Quechua	282	Huiliche	520	Haush/Maneken
145	Diaguita Ingamana	283	Pehuenche	530	Mak'a
146	Diaguita Quilmes	284	Picunche	540	Minuán
147	Diaguita Tolombón	285	Puelche	550	Ansilta
148	Diaguita Colastiné	290	Mapuche Tehuelche	560	Churumata
149	Diaguita Capayán	300	Mbya Guaraní	570	Jujuí
150	Diaguita Cacano	310	Moqoit/Mocoví	580	Michilingüe
160	Fiscara	320	Ocloya	970	No codificables
Guaraní		330	Omaguaca	980	No corresponde
171	Guaraní	340	Pilagá	990	Sin información

Pautas de codificación

Para el proceso de codificación asistida, se fijaron criterios a través de la definición de las siguientes pautas que unificaron las decisiones operativas del trabajo:

- Si el nombre del PI que aparece es reconocido como sinónimo de un nombre que ya existe en el nomenclador, se lo codifica con el código de ese pueblo y se lo añadirá como sinónimo en el NPI.
- Si aparece una respuesta que al codificar se reconoce como un PI pero no existe en el nomenclador, se lo codifica con "950", que funciona como código provisorio. Luego de la codificación, el equipo INAI-INDEC analiza todos esos casos e incorpora, cuando corresponda, el nombre en el NPI con un número de código propio. Lo mismo sucede con las especificidades de los pueblos.
- Si la respuesta no se corresponde con ninguna de las pautas anteriores y todavía hay duda, se deja sin codificar. Estos casos, luego de la codificación, serán analizados por el equipo INAI-INDEC para decidir su codificación según corresponda.

- d. Si en la respuesta aparece el nombre de dos pueblos distintos, se codifica por el nombre que esté escrito en primer término. Ejemplo: Toba Mataco se codifica por Toba; Ona Mapuche, por Ona. La excepción a esta pauta es que alguna combinación de dos nombres figure específicamente en el nomenclador. Por ejemplo: Mapuche Tehuelche, Kolla Quechua. Esta decisión metodológica es análoga a la tomada por el INAI al momento de organizar la elección de representantes ante el Consejo de Participación Indígena²⁶.
- e. Las declaraciones de nombres de pueblos que se acompañan con el nombre de algún país limítrofe y que, a su vez, son pueblos que tienen presencia en el territorio argentino, se los codifica con el código de ese pueblo. Ejemplo: "Aymara de Bolivia", será codificado como Aymara.

²⁶ "Las comunidades que se autorreconozcan pertenecientes a más de un pueblo originario, es decir[,] pueblo de pertenencia compuesto, se tomará en cuenta[,] a los fines de elegir su representación, el primero de los pueblos registrados en la Resolución de inscripción de su personería jurídica" (INAI, 2014, p. 8).

4. Codificación de Rama de actividad económica

En esta sección, se describen las características principales del proceso de codificación de la información proveniente de la pregunta abierta de rama de actividad económica (P34), realizada en el Censo Nacional de Población, Hogares y Viviendas 2022 (Censo 2022), en sus dos modalidades de relevamiento: **censo digital** y **censo por barrido territorial**.

Como se explica en la sección 1.1 de este informe, la CCN utiliza el SiCI, que habilita dos instancias de codificación: en primer lugar, la automática y luego, la asistida. Esta segunda instancia es una codificación realizada a través del análisis interpretativo de un equipo de codificación que utiliza:

- a) terminales informáticas programadas específicamente para el trabajo simultáneo en red y que presentan ayudas y filtros que asisten al codificador;
- b) protocolos, documentos y técnicas metodológicas constituidos a efectos de la codificación;
- c) conocimientos adquiridos mediante capacitación en los clasificadores utilizados y en el funcionamiento de las pantallas de codificación; y
- d) asistencia del personal técnico que tiene una amplia trayectoria en los procesos involucrados para la supervisión in situ durante el proceso de codificación.

A continuación, se presentan el detalle sobre las fases previas al procesamiento de la información, las estrategias y decisiones metodológicas planteadas para la codificación de la rama de actividad, la puesta en marcha, el análisis y el tratamiento de la información y, por último, las conclusiones sobre el trabajo desarrollado.

4.1 Fases previas al proceso de codificación

4.1.1 Consideraciones preliminares

En el Censo 2022, el bloque económico quedó conformado por las preguntas orientadas a medir la condición de actividad de las personas (P26 a P29 del cuestionario de Población), por la categoría ocupacional (P30) y por la rama de actividad económica del establecimiento (P34). El bloque incorporó como novedad una pregunta cerrada (P33), solicitada y elaborada por la Dirección Nacional de Cuentas Nacionales, que indagó sobre la actividad económica de los establecimientos a partir de 11 categorías construidas en base a la agregación de los 21 sectores principales que componen el Clasificador Nacional de Actividades Económicas (ClnAE 2010). Entre esas 11 categorías, 6 se corresponden con sectores definidos en ClnAE 2010²⁷ a nivel de letra y 5 resultan del agrupamiento o desdoblamiento de sectores del clasificador.

²⁷ Los sectores que se relacionan en forma directa con el ClnAE son: 3) comercio; 4) industria; 6) construcción; 7) transporte o almacenamiento; 8) hotel o restaurante; 9) banco, financiera o aseguradora.

El bloque económico del formulario quedó del siguiente modo:

PARA TODAS LAS PERSONAS DE 14 AÑOS Y MÁS																									
26 Durante la semana pasada ¿trabajó por lo menos una hora, sin contar las tareas domésticas de su hogar? Sí ☐ 1 → Pase a 30 No ☐ 2	31 En ese trabajo ¿le descuentan para la jubilación? Sí ☐ 1 → Pase a 32 No ☐ 2 Ignorado ☐ 9																								
27 En esa semana ¿hizo alguna change, fabricó algo para vender afuera, ayudó a un familiar o amigo en su chacra o negocio? Sí ☐ 1 → Pase a 30 No ☐ 2	32 En ese trabajo ¿aporta por sí mismo para la jubilación? Sí ☐ 1 No ☐ 2 Ignorado ☐ 9																								
28 En esa semana ¿tenía trabajo y no concurrió? No concurrió por: vacaciones, enfermedad, suspensión con pago, conflicto laboral, mal tiempo, etcétera. Sí ☐ 1 → Pase a 30 No ☐ 2	33 La actividad principal de la empresa, negocio, institución en la que trabaja o del trabajo que realiza por su cuenta ¿es... Refiere a la actividad del lugar donde trabaja la persona censada. <table style="width: 100%; border: none;"> <tr><td>educación o salud privada?</td><td style="text-align: right;">☐ 1</td></tr> <tr><td>administración pública, educación o salud pública?</td><td style="text-align: right;">☐ 2</td></tr> <tr><td>comercio?</td><td style="text-align: right;">☐ 3</td></tr> <tr><td>industria?</td><td style="text-align: right;">☐ 4</td></tr> <tr><td>agropecuaria, pesca o minería?</td><td style="text-align: right;">☐ 5</td></tr> <tr><td>construcción?</td><td style="text-align: right;">☐ 6</td></tr> <tr><td>transporte o almacenamiento?</td><td style="text-align: right;">☐ 7</td></tr> <tr><td>hotel o restaurante?</td><td style="text-align: right;">☐ 8</td></tr> <tr><td>banco, financiero o aseguradora?</td><td style="text-align: right;">☐ 9</td></tr> <tr><td>empresa de electricidad, gas o agua?</td><td style="text-align: right;">☐ 10</td></tr> <tr><td>otros servicios personales, técnicos o profesionales?</td><td style="text-align: right;">☐ 11</td></tr> <tr><td>Ignorado</td><td style="text-align: right;">☐ 99</td></tr> </table>	educación o salud privada?	☐ 1	administración pública, educación o salud pública?	☐ 2	comercio?	☐ 3	industria?	☐ 4	agropecuaria, pesca o minería?	☐ 5	construcción?	☐ 6	transporte o almacenamiento?	☐ 7	hotel o restaurante?	☐ 8	banco, financiero o aseguradora?	☐ 9	empresa de electricidad, gas o agua?	☐ 10	otros servicios personales, técnicos o profesionales?	☐ 11	Ignorado	☐ 99
educación o salud privada?	☐ 1																								
administración pública, educación o salud pública?	☐ 2																								
comercio?	☐ 3																								
industria?	☐ 4																								
agropecuaria, pesca o minería?	☐ 5																								
construcción?	☐ 6																								
transporte o almacenamiento?	☐ 7																								
hotel o restaurante?	☐ 8																								
banco, financiero o aseguradora?	☐ 9																								
empresa de electricidad, gas o agua?	☐ 10																								
otros servicios personales, técnicos o profesionales?	☐ 11																								
Ignorado	☐ 99																								
29 Durante las últimas cuatro semanas ¿buscó trabajo de alguna manera? Contestó avisos de diario o internet, consultó parientes u otras personas, puso carteles, envió curriculum, hizo algo para ponerse por su cuenta, etcétera. Sí ☐ 1 } Si es mujer de 14 a 49 años pase a 35. No ☐ 2 } Si no, fin de la entrevista para esta persona.	34 ¿Cómo describiría en detalle esa actividad principal de la empresa, negocio, institución en la que trabaja o del trabajo que realiza por su cuenta?																								
30 Ese trabajo ¿lo hace como... Si tiene más de un trabajo, refiere al que le lleva más horas. <table style="width: 100%; border: none;"> <tr><td>servicio doméstico?</td><td style="text-align: right;">☐ 1</td></tr> <tr><td>empleada(o) u obrera(o)?</td><td style="text-align: right;">☐ 2</td></tr> <tr><td>cuenta propia?</td><td style="text-align: right;">☐ 3</td></tr> <tr><td>patrón(a) o empleador(a)?</td><td style="text-align: right;">☐ 4</td></tr> <tr><td>trabajador(a) familiar?</td><td style="text-align: right;">☐ 5</td></tr> <tr><td>Ignorado</td><td style="text-align: right;">☐ 9</td></tr> </table> } → Pase a 32	servicio doméstico?	☐ 1	empleada(o) u obrera(o)?	☐ 2	cuenta propia?	☐ 3	patrón(a) o empleador(a)?	☐ 4	trabajador(a) familiar?	☐ 5	Ignorado	☐ 9													
servicio doméstico?	☐ 1																								
empleada(o) u obrera(o)?	☐ 2																								
cuenta propia?	☐ 3																								
patrón(a) o empleador(a)?	☐ 4																								
trabajador(a) familiar?	☐ 5																								
Ignorado	☐ 9																								

El Censo 2022 fue el primero en el que no se preguntó por la ocupación de las personas. Desde el Primer Censo de la República Argentina, esta pregunta estuvo presente en todos los operativos censales. Además, en algunas de las últimas experiencias, también se midió la condición de actividad, la categoría ocupacional, el tamaño del establecimiento donde se desarrolla la actividad, el sector del establecimiento (estatal o privado), la rama de actividad económica de los establecimientos; preguntas que, junto con la de la ocupación de las personas, se consideró que daba una buena medición del bloque económico²⁸.

En las reuniones de trabajo del Censo, durante la etapa de diseño del operativo censal, la CCN recomendó replicar la batería de preguntas para el bloque económico ya adoptadas en otros operativos censales y de encuesta (es decir, incluyó tamaño y sector del establecimiento económico y datos de la ocupación), argumentando criterios metodológicos sobre el proceso de clasificación. Tal recomendación devino de la experiencia y de la realización de pruebas²⁹ que mostraron dificultades al momento del relevamiento de la información, tales como la posible confusión entre los conceptos de ocupación y actividad económica: según lo registrado, es usual que una parte de la población responda estas preguntas e invierta el orden de las respuestas. Por lo tanto, al haber dos preguntas, una parte respondería con su ocupación en lugar de brindar información sobre la actividad económica del establecimiento donde trabaja. Esto disminuiría la calidad de la información y de los datos obtenidos, y afectaría la comparabilidad con relevamientos anteriores, en particular con la omisión de los datos de ocupación.

Sin embargo, el equipo a cargo de la coordinación del Censo descartó esta opción en dichas reuniones. Los motivos aducidos fueron el espacio que ocuparían estas preguntas en la cédula censal, el tiempo extra de relevamiento que supondrían y el costo que implicaría la posterior codificación de la pregunta de ocupación.³⁰

Las características adoptadas por el diseño final del bloque económico del Censo 2022 requirieron la readecuación de las estrategias de procesamiento y codificación de la pregunta abierta de rama de actividad. Así, además de la labor iniciada en el año 2017 para la planificación del trabajo, a lo largo de 2021 y el inicio de 2022, la CCN trabajó en esta readecuación.

²⁸ En el Anexo 3, se muestra el diseño del bloque económico del formulario del Censo 2001 a modo de ilustración de lo aquí descripto.

²⁹ Por ejemplo, se realizaron pruebas de codificación con bases de datos de la EPH para el período 2016-2020.

³⁰ Incluso, se descartó la posibilidad, propuesta por la CCN, de relevar la pregunta de ocupación sin codificarla para no demorar la publicación de los datos censales y evitar este costo de procesamiento, pero permitir que la información quedara disponible.

4.1.2 Instrumentos para la clasificación y codificación

Para el operativo censal, la Dirección Técnica del INDEC solicitó a la CCN la utilización del Clasificador Nacional de Actividades Económicas (ClaNAE) en reemplazo de la Clasificación de Actividades Económicas para Encuestas Sociodemográficas (CAES Mercosur) utilizada por el Instituto para la codificación de relevamientos a hogares. Ambos instrumentos son adaptaciones, una regional (CAES) y otra nacional (ClaNAE), de la Clasificación Industrial Internacional Uniforme de todas las actividades económicas Revisión 4 (CIIU Rev. 4), elaborada por Naciones Unidas.

La CAES tiene un nivel de desagregación menor que la CIIU. Por su parte, el ClaNAE posee un nivel de desagregación mayor que la clasificación internacional, expresado en una apertura máxima de codificación de hasta 5 dígitos (la CIIU tiene 4). Esta característica del ClaNAE se fundamenta en la naturaleza del nomenclador, orientada a la medición y clasificación en relevamientos realizados en las mismas unidades económicas, por ende, dispone a tal efecto categorías de subclases consideradas relevantes en la estructura económica local.³¹

La decisión de reemplazar el uso de la CAES por el ClaNAE en el Censo 2022 implicó la modificación de parte de los procedimientos y herramientas de codificación disponibles en el Instituto. Hubo que reelaborar el curso y manual de codificación. Se trabajó en la reasignación de códigos y corrección de frases, que se incorporaron posteriormente al SiCI para garantizar mejoras en la codificación automática. Se trabajó en la adaptación del sistema informático que garantice una codificación que unifique criterios de interpretación, disminuya los tiempos de codificación y permita una rápida recodificación en caso de necesitar cambiar el criterio asignado.

4.1.3 El Sistema de Codificación Informatizada y la codificación asistida. Su preparación

La adecuación e implementación del SiCI supuso la creación de un nuevo módulo, adaptado a la codificación del ClaNAE³².

El SiCI cuenta con un diccionario de codificación de rama de actividad que consiste en un listado de frases con sus respectivos códigos de la CAES³³ y que se usa en los relevamientos a hogares. La idea fue asignar a estas frases un código del ClaNAE, para utilizarlo en la codificación del Censo 2022. Entre julio y septiembre de 2021, la CCN recodificó manualmente las bases más actuales disponibles de la Encuesta Permanente de Hogares (2016-2020). Dada la magnitud de la tarea, se priorizaron los casos que tenían mayor frecuencia, y se asignó un código del ClaNAE a más de 50.000 frases.³⁴

También se desarrolló un diccionario de codificación con el ClaNAE para el Censo Nacional Económico (CNE) 2020-2021. Se tomaron de aquí los casos que tuvieran la mayor probabilidad de parecerse a la información proveniente de una encuesta a hogares, se revisaron y se sumaron al nuevo diccionario de codificación del ClaNAE para el Censo 2022.

4.2 Proceso de codificación censal

La información para codificar no ingresó de una sola vez al sistema, sino que se la fraccionó por lotes: por modalidad de relevamiento y por provincia completa o parcial si el volumen de información superaba determinado valor óptimo de trabajo. Una vez acondicionada la información y los datos para procesar, se usó el SiCI para codificar automáticamente los casos que coincidían con el diccionario. Los que no se pudieron codificar de esta manera, se destinaron a la codificación asistida. Cada lote codificado sirvió para la retroalimentación del diccionario, que resultó en un aumento de los casos procesados de manera automática, y para la corrección y actualización de los instrumentos de codificación.³⁵

³¹ Por esta razón, la CCN sugirió en principio no utilizar el ClaNAE para el relevamiento de los hogares.

³² Ver apartado 1.1 El Sistema de Codificación Informatizada (SiCI).

³³ El método que se utiliza en esta codificación es el de autofrase.

³⁴ El código de la CAES asignado originalmente en los casos de estas bases de datos sirvió como guía para acelerar la nueva asignación, dado que ambos clasificadores de rama de actividad comparten el primer y, muchas veces, el segundo dígito.

³⁵ Puede verse este proceso en forma de diagrama en el Anexo 4.

Por otra parte, al avanzar la codificación y notarse la cantidad de casos en que se declaraba la ocupación de la persona censada en lugar de la rama de actividad, se implementó una pauta complementaria que consistió en crear un nuevo diccionario de aplicación condicional, con nombres de ocupaciones que, en determinadas circunstancias, permitía una asignación de un código de actividad con un grado razonable de certidumbre.³⁶

4.2.1 Organización de la capacitación del equipo de codificación

Para realizar la codificación asistida primero se capacitó a un equipo externo que estaría a cargo de codificar. En agosto de 2021, comenzó la organización del curso para la codificación del Censo Experimental.³⁷ Se dispuso la capacitación para 30 personas, con el propósito de que esta instancia oficiara como un primer acercamiento a la definitiva, pero como hubo pocos casos que brindasen información en la pregunta de actividad económica³⁸, no fue necesario el empleo del equipo.

Entre el 20 y el 22 de diciembre de 2021, la CCN procesó y codificó estos datos. El principal hallazgo fue el alto porcentaje de respuestas orientadas a la descripción de la ocupación de las personas en lugar de las actividades económicas de los establecimientos, tal como se había advertido previamente a la realización del Censo Experimental.³⁹

En función de lo anterior, se recuperó el material que se había preparado originalmente para la codificación del Censo Experimental y se trabajó en su readaptación para la capacitación de las 80 personas encargadas de la codificación definitiva.

Finalmente, los documentos y herramientas seleccionados fueron:

- El Clasificador Nacional de Actividades Económicas 2010, revisión 2017 (ClaNAE 2010), con sus explicaciones y notas metodológicas, estructura en árbol desplegable a sección (1 dígito) y división (2 dígitos).
- Manual de Capacitación: manual de codificación de actividad del ClaNAE 2010.
- Documentación y material de exposición explicativos.
- Casos de ejercitación para la evaluación individual y colectiva de los contenidos.

4.2.1.1 La capacitación

Entre el 16 y el 19 de agosto de 2022, la CCN capacitó a 80 personas en la sede central de la Universidad Nacional de La Matanza (UNLaM). El objetivo fue proporcionarles las herramientas conceptuales y operativas necesarias para la realización de un análisis adecuado de las preguntas y respuestas sobre la rama de actividad económica para su posterior codificación con el ClaNAE 2010. Para ello, el curso apuntó a la profundización y práctica en el uso del clasificador, en las técnicas de codificación y en el uso del sistema informático.

La infraestructura de trabajo brindada por la Universidad, tanto durante la capacitación como a lo largo del operativo de codificación, constó de un aula de uso exclusivo con una superficie de 50 m² y mobiliario de trabajo, donde se dispusieron 40 terminales de computación y un servidor. Esto conformaba una red interna que trabajaba con los protocolos de seguridad del Instituto para garantizar la confidencialidad de la información. También había dos computadoras para realizar el control de calidad del proceso de codificación.

A su vez, el personal de la CCN a cargo de la supervisión dispuso de tres computadoras conectadas a la red interna de codificación y a las terminales de control de calidad. El equipamiento informático fue provisto e instalado por el INDEC.

³⁶ Esta aplicación condicional supone lo que en la terminología del SiCI se denomina “microproceso”: la asignación de un código no es inmediata, sino que depende de que previamente se realicen una serie de pasos de verificación.

³⁷ Aprovechando las distintas instancias de pruebas preparatorias del Censo y la información de otros relevamientos, se realizaron varias mediciones que incluyeron: ejercicios de codificación con el ClaNAE sobre información de la Encuesta Permanente de Hogares, utilización de la información provista por el Sistema Integrado Previsional Argentino (SIPA) para cotejarla con la de las pruebas piloto del Censo, formulación de distintas baterías de preguntas del bloque económico en dichas pruebas, etcétera.

³⁸ Fueron 4.782 casos, de los cuales 3.300 se procesaron de manera automática.

³⁹ Ver más arriba, en el apartado 4.1.1 Consideraciones preliminares, el párrafo que sigue a las imágenes de los formularios censales.

Debido a las características del espacio asignado para el proceso de codificación, se organizó a las y los codificadores en dos turnos de trabajo, mañana y tarde (grupo turno mañana, de 10 a 15 h; grupo turno tarde, de 15 a 20 h). Se les asignó un nombre de usuario: CODIFICADOR seguido del número de orden (empezando por el CODIFICADOR01 hasta llegar al CODIFICADOR80) para definir la clave primaria de identificación.

4.2.2 Secuencia general en la codificación de la rama de actividad⁴⁰

4.2.2.1 Codificación de la pregunta 34

Para la codificación de la pregunta 34 del cuestionario censal (pregunta abierta sobre la rama de actividad económica) se definieron y adoptaron los siguientes criterios y pasos.

- 1) Para todos los casos en que P30 tenía valor “1” (*Servicio doméstico*), se asignó el código de actividad CODACT⁴¹ “97000” (*Actividades de los hogares como empleadores de personal doméstico*). Es decir, se revisó la pregunta 30, que indaga sobre la categoría ocupacional, y los casos con valor “1” quedaron codificados automáticamente⁴².
- 2) Se ejecutó el SiCI, lo que resultó en la asignación de códigos de manera automática.
- 3) El resto de los registros pasaron a la línea de codificación asistida. Se preparó la información para codificarla en red y el equipo de codificación resolvió los casos que les iban apareciendo en pantalla. Este fue el paso que demandó la mayor cantidad de tiempo y de recursos asignados ya que involucró una dotación permanente de personal para codificar y supervisar.
- 4) Se procedió a una revisión de aquellos casos con código V (casos vacíos, sin información declarada) y Z (información insuficiente para codificar, valor asignado en la codificación automática o asistida), tomando como guía la información declarada en la pregunta 33. Como se señaló al inicio, esta es una pregunta con opciones de respuesta predefinidas, referida a la actividad del lugar en el que trabaja la persona censada. El propósito fue recuperar información realizando las siguientes imputaciones:

Si CODACT = V o Z y si P33 = **3** → CODACT = **G**

Si CODACT = V o Z y si P33 = **4** → CODACT = **C**

Si CODACT = V o Z y si P33 = **6** → CODACT = **F**

Si CODACT = V o Z y si P33 = **7** → CODACT = **H**

Si CODACT = V o Z y si P33 = **8** → CODACT = **I**

Si CODACT = V o Z y si P33 = **9** → CODACT = **K**⁴³

Hasta aquí los pasos referidos a la codificación de la pregunta 34.

4.2.2.2 Tratamiento de la pregunta 33

Posteriormente, se revisó y adecuó la información de la pregunta 33 para que sea consistente con la información de la pregunta 34⁴⁴ en dos pasos.

⁴⁰ En el anexo 4, se encuentra un esquema de lo que se desarrolla en esta sección.

⁴¹ CODACT es el nombre del campo en donde se ubica el valor codificado de la variable “Rama de actividad económica” relevada en la pregunta 34.

⁴² Para tomar esta decisión, se revisaron estos casos por si la información de la pregunta abierta remitía a una actividad distinta. En la inmensa mayoría de las veces la asignación era correcta, por lo que se aceptó un pequeño margen de error en aras de disminuir el trabajo de línea y acelerar la codificación. A este conjunto de registros no se les aplicaron los pasos posteriores y quedó como código final 97000.

⁴³ Para los casos de CODACT = V o Z y algún valor de P33 distinto a los mencionados aquí, se mantuvieron los códigos V o Z ya que no hay manera de recuperar más información dado que se trata de agrupamientos de sectores que no se pueden separar a nivel de sección.

⁴⁴ Se priorizó el valor obtenido de los textos declarados por las personas censadas (pregunta 34) por encima del valor dado en las respuestas predefinidas (pregunta 33), ya que en estas últimas se verificaron algunas inconsistencias o falta de comprensión de las opciones propuestas para responder esta pregunta. En los casos en que la información declarada era consistente en las dos preguntas y cuando se usaba la pregunta cerrada para asignar el valor de la pregunta abierta (casos V y Z), el valor original de la pregunta 33 se mantuvo.

1) Se hizo la conversión de la información codificada de rama de actividad (pregunta N° 34) a su valor de letra ClaNAE⁴⁵. Por ejemplo, si la pregunta 34 fue codificada como “45” (*Venta, mantenimiento y reparación de vehículos automotores y motocicletas*), o como “463” (*Venta por mayor de alimentos, bebidas y tabaco*) o como “47741” (*Venta por menor de artículos de óptica y fotografía*), la sección ClaNAE que corresponde en todos estos casos es “G” (*Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas*)⁴⁶.

2) Se realizó la conversión de la sección ClaNAE al valor correspondiente de la pregunta 33, para así obtener el valor corregido de esta pregunta. Por ejemplo, a los casos que en sección ClaNAE tenían asignada la “G”, se les asignó el valor “3” (*Comercio*) en la pregunta 33 corregida.⁴⁷

4.2.2.3 Entrega de la información para cierre y consistencias

Luego de codificar la pregunta 34 y de corregir la pregunta 33, la CCN entregó la información a la DNESP y se sugirió revisar, tal como se había acordado, la consistencia y corrección de las preguntas cerradas del bloque económico⁴⁸.

Además de depurar la información y los datos, la DNESP pudo realizar las siguientes consistencias:

a) En cuanto a la condición de actividad (establecida mediante las preguntas 26 a 29), debían cambiar el estado de activos y considerarse inactivos los casos en que la pregunta 34 fuera W; es decir, la sección ClaNAE correspondiente a las personas no económicamente activas (personas jubiladas, estudiantes, amas de casa, etcétera).

b) En cuanto a la categoría ocupacional, debía modificar y asignar valor “1” (*Servicio doméstico*) a la pregunta 30 en los casos en que esta tuviese otro valor, pero la pregunta 34 fuera “97000”.

Al concluir estos pasos, se contó con la información definitiva para la publicación⁴⁹ y para el análisis de las características económicas de la población.

4.2.3 Estrategias y decisiones metodológicas

Antes y durante del operativo, la CCN tomó y readecuó una serie de pautas y criterios para la codificación de la pregunta de rama de actividad económica. Las decisiones iniciales se basaron en la revisión de la codificación de otros operativos y en las pruebas piloto y Censo Experimental. Una vez que comenzó el análisis y procesamiento de la información del Censo, se realizaron controles de calidad de la codificación automática y asistida que, entre otras cosas, aportaron datos para evaluar los materiales de codificación. Así, a lo largo del operativo, se sumaron nuevas pautas de trabajo si se consideraba necesario.

El criterio general más relevante dispuso que el código asignado se estableciera en base a la descripción de la actividad económica y que, cuando la información no fuera suficiente para codificar al máximo nivel de apertura del clasificador (5 dígitos), se debiera codificar a menos dígitos (o, en última instancia, a sección) para no perder la información sobre el caso⁵⁰.

Por otra parte, el análisis del Censo Experimental llevó también a prever una característica adicional, que luego se constató, sobre el tipo de información producido por el operativo censal: una proporción considerable de respuestas basadas en datos sobre ocupación en lugar de declaraciones sobre rama de actividad económica.⁵¹ Por esta razón,

⁴⁵ La sección ClaNAE es el valor más agregado del Clasificador y puede funcionar como variable independiente en algunos tabulados. El ClaNAE tiene una estructura jerárquica con cinco niveles de apertura: el primer nivel, correspondiente al sector (codificación a letra); el segundo, a la división de la actividad (codificación a dos dígitos); el tercero, al grupo de dicha actividad (codificación a tres dígitos); el cuarto, a la clase de dicha rama de actividad (codificación a cuatro dígitos); y, por último, el quinto nivel responde a la subclase de la actividad declarada (codificación a cinco dígitos).

⁴⁶ La tabla de conversión de los valores de la pregunta 34 a la sección ClaNAE se encuentra en el anexo 6.

⁴⁷ La tabla de conversión de los valores de la sección ClaNAE a la pregunta 33 corregida se encuentra en el anexo 7.

⁴⁸ La CCN trabajó con bases sin consistir y sin jerarquizar y, por lo tanto, incluyó casos que la DNESP podría quitar del universo al revisar los datos finales.

⁴⁹ Disponible en: https://censo.gob.ar/wpcontent/uploads/2024/02/censo2022_caracteristicas_economicas.pdf

⁵⁰ Luego del análisis de la información proveniente de las pruebas censales, la CCN previó la recepción de un volumen importante de respuestas con información insuficiente para la clasificación de rama de actividad al máximo nivel de apertura, aspecto que incidió en la decisión metodológica adoptada.

⁵¹ Esto se puede observar más adelante en el cuadro 16.

se estableció que en los casos en que se declaraba una actividad independiente o por cuenta propia en la pregunta 30, la información provista se podría utilizar para codificar la rama de actividad.⁵² Por el contrario, si la respuesta correspondía a la declaración de un trabajo realizado de manera asalariada, la información se consideraría no codificable porque se desconocía por completo la rama del establecimiento en el que se desarrollaba la actividad.

Un ejemplo que ilustra este criterio: si en la pregunta abierta referida a la actividad principal la respuesta fue “médico” y la categoría ocupacional era “asalariado”, al caso se lo consideró no codificable (Z). Un médico puede ser asalariado en distintos establecimientos, como por ejemplo un establecimiento fabril, en el que la actividad económica no se enmarca en la atención de la salud sino en la de la fábrica en la que esta persona está empleada. En cambio, en un caso en que la respuesta fuese la misma, pero con el trabajo declarado como realizado por cuenta propia o como patrón, se asignó el código 86 (*Servicios de atención a la salud humana*)⁵³.

A los casos que brindasen información considerada insuficiente para ser codificada, se les asignó la letra Z. Este criterio se aplicó tanto a las respuestas en las que no se pudo diferenciar el lugar de realización de la actividad, por ejemplo, una respuesta donde solamente se declaraba “abogada” y la persona era asalariada, como a las que la descripción de la actividad fue escueta o imprecisa para esta variable, por ejemplo, “sector de producción”, o que solo se describieron con palabras que resultaron insuficientes para determinar el sector económico involucrado, por ejemplo, “muebles”, que podría ser tanto su producción como su venta.

Además de los casos no codificables, se observaron respuestas que remitían a la población no económicamente activa, por ejemplo, quienes realizan tareas de trabajo doméstico no remunerado, estudiantes y personas jubiladas. En estos casos, se codificó con la letra W.

Los casos sin información (casos vacíos), se codificaron con la letra V.

Estas tres letras (V, W y Z) no forman parte del clasificador, ya que no expresan ramas de actividad. Sin embargo, se incorporaron por su importancia para el proceso de clasificación de los resultados.

Una vez establecidas las pautas de trabajo, desde el 22 de agosto de 2022 hasta enero de 2023, se llevó a cabo el operativo de codificación asistida de la información del censo digital.

Las dificultades en el escaneo de los formularios censales del barrido territorial tuvo retrasos en el envío de la información y los datos correspondientes, además de problemas de calidad: la información que surgió de la lectura y escaneo de las cédulas censales trajo datos difíciles de interpretar. Por ende, se rechazó y se inició un proceso de relectura de la información que continuó hasta que se logró que la información tuviera los estándares establecidos para considerarla válida⁵⁴. Esto obligó a suspender la codificación durante el mes de febrero, hasta que se volviera a tener suficiente información para continuar con los procesos. Durante la etapa de planificación, la CCN había estimado una duración total de 270 días (9 meses) para la codificación asistida, pero por el motivo señalado la fecha de finalización se extendió a 10 meses.⁵⁵

Así, el segundo período de codificación asistida comenzó en marzo y culminó en junio. Durante este período, se codificó la información del Censo por barrido territorial y algunos datos residuales del Censo Digital.

4.3 Resultados del proceso de codificación

Los siguientes datos corresponden a la codificación de bases sin consistir debido a que la CCN acordó iniciar el proceso de esta manera por solicitud de la DNESP. Por esto, no se trata de los resultados definitivos del Censo, sino del proceso de codificación.

⁵² Se evaluó cada uno de los casos, ya que este criterio no podía aplicarse siempre con algún grado de certidumbre.

⁵³ La codificación, en el último ejemplo, fue a dos dígitos ya que la información que se brinda no es suficiente para una mayor desagregación.

⁵⁴ Además de lo requerido por la DNESP, la CCN estableció como umbral que el SiCI codificara automáticamente el 65% de los casos. Se tomaron como referencia los estándares de codificación automática de otros lotes en este mismo operativo.

⁵⁵ Al regresar al operativo de codificación asistida, ya en el mes de marzo, la recepción de las bases siguió presentando inconvenientes. Por este motivo y para no postergar el trabajo, la CCN decidió establecer un nuevo código y pauta, donde a los casos que presentaran información ilegible (casos que contaban con letras repetidas, con guiones, solo consonantes, etcétera) se les asignó la letra X. El propósito de este código, una vez reenviadas las bases por la empresa con sus respectivas correcciones, fue recuperar estos casos para volverlos a codificar.

En total se codificaron 20.428.615, casos, de los cuales 12.510.503 corresponden a la modalidad de censo digital y 7.918.112 a la de censo por barrio territorial (cuadro 12). En el caso de la modalidad de censo digital, 75% de los casos se codificaron automáticamente y 25% de manera asistida (cuadro 13). En lo que respecta a la modalidad de censo por barrio territorial, el 77% se codificó automáticamente y el 23% restante corresponde a la codificación asistida (cuadro 14).

Como se puede observar en el cuadro 15, el 11,3% de los casos correspondió a la letra Z, 5,3% provenientes de la modalidad de censo digital y 6% del censo por barrio territorial. Los casos W representaron el 0,6% de la codificación, 0,1% del censo digital y 0,5% del censo por barrio territorial. Por último, los casos codificados con la letra V fueron el 7%, 6,9% de la modalidad de censo digital y 0,1% del censo por barrio territorial.

Finalmente, 22,5% de los casos correspondieron a respuestas con nombre de ocupación en lugar de rama de actividad económica, que en el caso de la modalidad del censo digital fue de 17,4%, mientras que para el censo por barrio territorial fue de 31,0% (cuadro 16).

Cuadro 12. Rama de actividad económica. Distribución de la población total codificada, por modalidad de relevamiento, según jurisdicción. Año 2022

Jurisdicción	Total de población codificada	Modalidad de relevamiento	
		Censo digital	Censo por barrio territorial
Total del país	20.428.615	12.510.503	7.918.112
Ciudad Autónoma de Buenos Aires	1.594.083	1.237.764	356.319
Buenos Aires	7.465.043	4.846.763	2.618.280
Catamarca	181.827	85.306	96.521
Chaco	424.503	211.754	212.749
Chubut	220.093	183.663	36.430
Córdoba	1.732.162	1.128.165	603.997
Corrientes	447.040	261.755	185.285
Entre Ríos	711.553	371.469	340.084
Formosa	241.564	116.343	125.221
Jujuy	330.911	140.252	190.659
La Pampa	181.050	119.557	61.493
La Rioja	165.142	82.644	82.498
Mendoza	863.084	533.725	329.359
Misiones	501.256	232.807	268.449
Neuquén	327.277	213.718	113.559
Río Negro	339.713	225.016	114.697
Salta	567.976	226.655	341.321
San Juan	311.026	156.616	154.410
San Luis	233.725	150.792	82.933
Santa Cruz	149.788	105.480	44.308
Santa Fe	1.544.432	1.037.555	506.877
Santiago del Estero	373.876	193.618	180.258
Tierra del Fuego, Antártida e Islas del Atlántico Sur	87.345	65.131	22.214
Tucumán	656.172	324.776	331.396
Jurisdicción no identificada ⁽¹⁾	777.974	259.179	518.795

⁽¹⁾ Casos recibidos por la CCN sin información de jurisdicción.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

Cuadro 13. Rama de actividad económica. Distribución de la población codificada en el censo digital, por modalidad de codificación, según jurisdicción. Año 2022

Jurisdicción	Total de población codificada	Modalidad de codificación	
		Codificación automática	Codificación asistida
Total del país	12.510.503	9.403.432	3.107.071
Ciudad Autónoma de Buenos Aires	1.237.764	890.736	347.028
Buenos Aires	4.846.763	3.459.011	1.387.752
Catamarca	85.306	68.521	16.785
Chaco	211.754	170.720	41.034
Chubut	183.663	140.191	43.472
Córdoba	1.128.165	877.470	250.695
Corrientes	261.755	215.018	46.737
Entre Ríos	371.469	296.590	74.879
Formosa	116.343	97.192	19.151
Jujuy	140.252	108.944	31.308
La Pampa	119.557	96.199	23.358
La Rioja	82.644	67.499	15.145
Mendoza	533.725	410.960	122.765
Misiones	232.807	187.073	45.734
Neuquén	213.718	160.974	52.744
Río Negro	225.016	171.890	53.126
Salta	226.655	178.033	48.622
San Juan	156.616	123.822	32.794
San Luis	150.792	120.089	30.703
Santa Cruz	105.480	83.100	22.380
Santa Fe	1.037.555	814.214	223.341
Santiago del Estero	193.618	160.040	33.578
Tierra del Fuego, Antártida e Islas del Atlántico Sur	65.131	50.215	14.916
Tucumán	324.776	256.585	68.191
Jurisdicción no identificada ⁽¹⁾	259.179	198.346	60.833

⁽¹⁾ Casos recibidos por la CCN sin información de jurisdicción.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

Cuadro 14. Rama de actividad económica. Distribución de la población codificada en el censo por barrido territorial, por modalidad de codificación, según jurisdicción. Año 2022

Jurisdicción	Total de población codificada	Modalidad de codificación	
		Codificación automática	Codificación asistida
Total del país	7.918.112	6.090.166	1.827.946
Ciudad Autónoma de Buenos Aires	356.319	259.355	96.964
Buenos Aires	2.618.280	2.076.603	541.677
Catamarca	96.521	69.862	26.659
Chaco	212.749	170.802	41.947
Chubut	36.430	19.521	16.909
Córdoba	603.997	454.817	149.180
Corrientes	185.285	143.017	42.268
Entre Ríos	340.084	280.810	59.274
Formosa	125.221	105.104	20.117
Jujuy	190.659	145.813	44.846
La Pampa	61.493	50.356	11.137
La Rioja	82.498	59.345	23.153
Mendoza	329.359	254.683	74.676
Misiones	268.449	213.242	55.207
Neuquén	113.559	88.261	25.298
Río Negro	114.697	82.615	32.082
Salta	341.321	233.509	107.812
San Juan	154.410	124.179	30.231
San Luis	82.933	57.374	25.559
Santa Cruz	44.308	34.774	9.534
Santa Fe	506.877	408.938	97.939
Santiago del Estero	180.258	127.045	53.213
Tierra del Fuego, Antártida e Islas del Atlántico Sur	22.214	16.689	5.525
Tucumán	331.396	235.035	96.361
Jurisdicción no identificada ⁽¹⁾	518.795	378.417	140.378

⁽¹⁾ Casos recibidos por la CCN sin información de jurisdicción.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

Cuadro 15. Rama de actividad económica. Distribución de casos de población con códigos complementarios del ClaNAE, según modalidad de relevamiento. Año 2022

Códigos complementarios de ClaNAE	Total	Porcentaje total ⁽¹⁾	Modalidad de relevamiento	
			Censo digital	Censo por barrido territorial
		%		
Población no económicamente activa (W)	132.287	0,6	22.841	109.446
Información insuficiente para codificar (Z)	2.313.294	11,5	1.077.451	1.235.843
Sin respuesta (V)	1.428.602	7,0	1.406.221	22.381

⁽¹⁾ Refiere al porcentaje sobre el total codificado.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

Cuadro 16. Distribución de respuestas con nombre de ocupación en lugar de rama de actividad económica, por modalidad de relevamiento, según jurisdicción. Año 2022

Jurisdicción	Porcentaje total de casos con nombre de ocupación	Modalidad de relevamiento	
		Censo digital	Censo por barrido territorial
		%	
Total del país ⁽¹⁾	22,5	17,4	31,0
Ciudad Autónoma de Buenos Aires	18,6	14,5	33,2
Buenos Aires	25,2	19,9	35,1
Catamarca	32,4	29,4	35,1
Chaco	22,2	11,1	33,4
Chubut	18,1	17,1	23,7
Córdoba	28,2	26,4	31,8
Corrientes	21,4	20,9	22,2
Entre Ríos	11,8	12,1	11,6
Formosa	13,5	8,2	18,5
Jujuy	20,2	9,2	28,3
La Pampa	17,7	9,7	33,4
La Rioja	23,8	13,2	34,6
Mendoza	17,2	11,7	21,1
Misiones	15,2	13,7	31,8
Neuquén	23,3	12,7	31,4
Río Negro	16,6	13,4	23,1
Salta	26,5	13,1	35,5
San Juan	13,8	8,5	19,3
San Luis	16,9	13,3	23,7
Santa Cruz	11,9	7,3	23,1
Santa Fe	7,8	16,3	32,2
Santiago del Estero	21,5	13,5	27,8
Tierra del Fuego, Antártida e Islas del Atlántico Sur	20,3	7,5	33,1
Tucumán	25,2	11,7	38,5

⁽¹⁾ No se incluyen aquí los casos de jurisdicción no identificada.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

4.4 Indicadores de seguimiento de la codificación asistida

Se establecieron indicadores desde el inicio del proceso de codificación para evaluar de forma constante la codificación asistida. En proyectos de esta magnitud, que implican la gestión de una cantidad significativa de información y con un plazo de ejecución definido, resulta imprescindible contar con indicadores cuantitativos consistentes que deben dar cuenta de la evolución del trabajo, los progresos o retrocesos, identificar posibles problemas y evaluar los ritmos laborales que garantizaren el cumplimiento de la tarea en el plazo acordado.

Se entendió que la semana era la unidad de tiempo que mejor ordenaba el seguimiento del trabajo, pues es el lapso apropiado para medir y observar la evolución de los indicadores⁵⁶:

- **la producción mensual:** la suma de lo codificado en un mes por el conjunto total de las y los codificadores;
- **la producción media diaria:** la totalidad de lo codificado por el conjunto de codificadores en un mes dividida por la cantidad de días de trabajo real de ese mes;
- **la productividad media diaria por codificador:** el promedio diario de lo codificado por codificador, considerando la cantidad de jornadas reales en las que estuvieron presentes⁵⁷.

El equipo de la CCN ingresaba diariamente en planillas la información de base con la que después se construían estos indicadores. El resultado para cada lapso que duró la codificación es el que se presenta en el cuadro siguiente:

Cuadro 17. Rama de actividad económica. Producción mensual, producción media diaria y producción media diaria, por codificador. Agosto 2022-mayo 2023

Año	Mes	Producción mensual	Producción media diaria	Producción media diaria por codificador
2022	Agosto ⁽¹⁾	89.308	9.923	128,3
	Septiembre	291.538	15.344	202,6
	Octubre	270.077	15.004	201,9
	Noviembre	367.646	15.318	213,6
	Diciembre	226.733	15.115	210,7
2023	Enero	222.304	13.076	228,0
	Marzo ⁽²⁾	386.048	20.318	297,4
	Abril	328.340	21.889	313,3
	Mayo	427.293	20.347	327,9

⁽¹⁾ Datos de dos semanas, ya que la codificación comenzó a mediados de agosto.

⁽²⁾ Durante febrero de 2023 no se codificó, por las razones mencionadas en el subapartado 4.2.3.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

Luego de la segunda quincena de agosto de 2022, cuando comenzó la labor y el grupo de codificación empezó a familiarizarse con su tarea, se observaron dos períodos diferenciados: desde septiembre hasta enero, a cuyo término se produce una pausa, y desde marzo hasta mayo.

En el primer período se tiene una producción media diaria estable, con una pequeña tendencia a la suba de la producción por codificador.

En el segundo período se observa un incremento, tanto en la producción media diaria, que se estabiliza en un valor más alto que el del primer período, como en la producción por persona codificadora, que asciende durante esos meses.

⁵⁶ Aquí se exponen los datos agrupados por unidad de tiempo mensual para no complejizar la presentación.

⁵⁷ Se tuvo en cuenta la cantidad real de codificadores en cada día para no distorsionar el promedio con ausencias eventuales.

Los factores que influyeron en estos valores fueron:

- El aprendizaje continuo y el incremento de la experiencia y la confianza del grupo de codificación en su labor.
- La capacidad adquirida para decidir respecto a los casos con información insuficiente para codificar (casos Z).
- La capacitación continua, prevista y suministrada por la CCN.
- La cantidad variable de codificadores a lo largo del proceso. Hubo rotación, reemplazos y licencias.
- El flujo irregular del ingreso de los cuestionarios escaneados al grupo de codificación.
- La calidad de la información recibida.

En conclusión, en términos generales, se verificaron los presupuestos que tenía la CCN en cuanto a un incremento de la producción y de la productividad con el avance del proceso de codificación.

4.5 El control de calidad

Se incluyó el control de calidad de los datos producidos en las tareas durante el proceso de codificación y tras su finalización. La metodología del control se diseñó en función de los siguientes objetivos:

- a) estimar el error en el proceso de codificación;
- b) descomponer y analizar el error total estimado según la modalidad de relevamiento (digital/barrido territorial) y según el tipo de codificación (automática/asistida), a nivel provincial y según codificador;
- c) identificar y tipificar posibles sesgos operativos o conceptuales que pudieron inducir el error clasificatorio a partir de una muestra de casos analizados;
- d) diseñar e implementar estrategias que tiendan a reducir el error clasificatorio durante el proceso de codificación censal para los siguientes lotes.

4.5.1 Estrategias de control de calidad

El análisis estadístico del error se abordó a partir de la selección de muestras probabilísticas⁵⁸ sobre la base de la información procesada y codificada mediante el SiCI, tanto de manera automática como asistida⁵⁹. La selección de las muestras de control se realizó por jurisdicción, por modalidad de codificación y por codificador para la codificación asistida. Esto posibilitó un análisis desagregado por dimensión geográfica y el diseño de estrategias dirigidas y focalizadas a la capacitación de recursos humanos en los casos que lo requirieron, a la corrección y a la ampliación de los diccionarios de codificación. En este último caso, el equipo de la CCN revisó las codificaciones, recomendó modificaciones en los diccionarios y sugirió nuevas autofrases⁶⁰.

A los dos meses del inicio de las tareas, se comenzó a evaluar la calidad de la codificación asistida para los casos relevados mediante el censo digital hasta que se terminó el procesamiento de esta modalidad. Luego se controló la calidad de los casos codificados correspondientes al relevamiento en terreno o censo por barrido territorial.

El método de pareo mediante recodificación independiente a doble ciego fue la metodología implementada para los controles de calidad de la codificación asistida. Mientras se iba procesando la información, se retiraban diariamente y de manera rotativa de la línea de codificación a dos personas, que pasaban a desarrollar la función de verificadores ("verificador 1" y "verificador 2"). La tarea consistía en la recodificación de la muestra seleccionada para el control de calidad.

⁵⁸ Elaboradas con asistencia de la Dirección de Metodología e Innovación Estadística y de la Coordinación de Investigación y Desarrollo Metodológico, ambas de la Dirección Nacional de Metodología e Infraestructura Estadística.

⁵⁹ Respecto de estos últimos, si bien en principio se usaron muestras del doble de tamaño, se decidió extraer una muestra de 750 casos en bases de hasta 100.000 registros (el máximo que se usó en la codificación en red).

⁶⁰ Ver la nota 40 al pie de página.

En ambos casos, trabajaron siguiendo los mismos criterios respecto a la codificación original y desconociendo el código asignado por los codificadores y la otra parte verificadora.

La metodología por doble ciego permite clasificar los casos de la muestra de control como lo indica el esquema 2:

Esquema 2. Clasificación de los casos de la muestra de control

Clasificación	Descripción	Decisión
A	Coincidencia entre el codificador y al menos un verificador, a dos dígitos de ClaNAE ⁶¹	Código correcto
B	Coincidencia entre los dos verificadores, a dos dígitos de ClaNAE, y divergencia del codificador	Error
C	Triple divergencia a dos dígitos de ClaNAE	El caso fue sometido a una instancia de metasupervisión por parte del equipo especializado de la CCN, que decidía la validez o error del código asignado por el codificador

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

4.5.2 Acciones implementadas para la reducción del error en la codificación asistida

Además de la supervisión permanente in situ del equipo técnico de la CCN para la asistencia técnica y resolución de dudas, durante el desarrollo de los trabajos de codificación asistida se implementaron recapitaciones periódicas y readecuaciones en el instrumental de codificación utilizado que consistieron en:

- **Recapitación periódica (grupal):** desde el inicio de las tareas de codificación, el equipo técnico de la CCN brindó capacitaciones y actualizaciones para cada uno de los grupos de codificación (turno mañana y turno tarde). Estas capacitaciones, que trabajaron en la resolución de prácticas seleccionadas a partir de una muestra cualitativa de casos, tuvieron por objeto clarificar determinadas dimensiones conceptuales del instrumento clasificatorio observadas como problemáticas, fortalecer el conocimiento sobre los protocolos y decisiones metodológicas y operativas adoptadas para la clasificación de la información, y resolver series específicas de respuestas al Censo identificadas como conflictivas.
- **Recapitación focalizada (individual):** tuvo por objetivo profundizar las competencias individuales a partir de la revisión de casos codificados con error clasificatorio.
- **Actualización permanente de los instrumentos de codificación:** a partir del surgimiento de dudas reiteradas, se establecieron pautas especiales para la codificación. El propósito de estas pautas fue estandarizar resoluciones y minimizar el error en caso de volver a codificar el mismo tipo de información.

4.5.3 Niveles de error estimados

A partir del análisis de la información producida por el proceso de control de calidad, se determinaron las siguientes tasas estimadas de error a dos dígitos del ClaNAE:

⁶¹ Dada la dispersión en los códigos asignados provocada por tratarse de un relevamiento a hogares y no a empresas, y a que los tabulados principales del Censo para rama de actividad económica se publican a dos dígitos, se tomó la decisión general de considerar el error en este nivel de desagregación de la variable.

Cuadro 18. Distribución del porcentaje de error estimado en la codificación de rama de actividad del censo, según modalidad de relevamiento y de codificación. Año 2022

Jurisdicción	Porcentaje de error total	Censo digital		Censo por barrido territorial	
		Porcentaje de error en la codificación automática	Porcentaje de error en la codificación asistida	Porcentaje de error en la codificación automática	Porcentaje de error en la codificación asistida
Total del país ⁽¹⁾	6,3	2,2	16,4	3,4	18,2
Ciudad Autónoma de Buenos Aires	5,9	1,7	15,7	2,9	21,5
Buenos Aires	6,1	2,1	15,8	3,3	19,0
Catamarca	6,8	1,4	18,9	3,4	23,1
Chaco	7,7	6,0	19,5	3,6	21,3
Chubut	6,6	2,3	18,3	1,6	19,6
Córdoba	6,0	2,3	18,0	1,2	23,3
Corrientes	5,5	1,3	20,2	2,9	21,8
Entre Ríos	3,1	1,0	19,2	1,0	3,6
Formosa	5,8	3,4	16,3	3,5	20,3
Jujuy	6,7	2,3	16,5	4,9	18,3
La Pampa	7,1	2,1	20,8	7,8	20,6
La Rioja	9,1	3,1	15,6	7,9	26,7
Mendoza	5,6	1,8	15,5	1,8	24,1
Misiones	8,6	4,9	18,6	7,9	16,8
Neuquén	7,7	2,5	16,1	7,8	24,8
Río Negro	7,9	2,6	15,1	9,5	20,7
Salta	6,3	2,8	15,3	1,0	19,8
San Juan	6,6	4,3	16,4	3,3	19,6
San Luis	9,2	7,4	16,1	10,6	7,9
Santa Cruz	7,6	4,1	17,5	3,6	30,9
Santa Fe	6,4	2,0	17,0	5,7	21,8
Santiago del Estero	8,8	5,0	16,7	5,2	24,0
Tierra del Fuego, Antártida e Islas del Atlántico Sur	7,4	3,8	19,4	3,4	24,5
Tucumán	7,2	2,5	18,7	5,9	15,1

⁽¹⁾ No se incluyen aquí los casos de jurisdicción no identificada.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

El error total estimado de codificación de rama de actividad se determinó en 6,31%.

Las diferencias entre las tasas de error obtenidas mediante el SiCI de forma automática y las de la codificación asistida se explican mayormente por diferencias en el tipo de información de los casos codificados mediante cada una de esas técnicas. Los casos para codificar en forma asistida son los más difíciles de interpretar, ya que han llegado a dicha instancia cuando el SiCI no pudo determinar un código automáticamente, y son los de poca frecuencia, que por eso no figuran en el diccionario de codificación⁶².

⁶² Se observa la misma tendencia en otras experiencias de codificación conducidas por el INDEC.

4.5.4 Fuentes de error identificadas

Aunque el análisis de los errores pueda requerir una indagación más detenida, se formulan aquí las siguientes observaciones sobre los factores identificados que podrían haber contribuido al error clasificatorio:

- El uso de clasificadores económicos para codificar declaraciones producidas en relevamientos a hogares. Si se acude a una empresa para relevar la actividad económica realizada, se suele contar con un informante clave que conoce cuál es. En el Censo, la persona respondiente no siempre está al tanto del conjunto de actividades realizadas por la empresa; o pudo haber respondido otro miembro del hogar que conoce menos aún esa actividad.
- La incidencia de declaraciones con información sobre tareas u ocupaciones de las personas censadas, en lugar de información sobre la actividad económica.
- La calidad de la digitalización de las cédulas censales.

4.6 Conclusiones

El Censo 2022 introdujo diversas innovaciones, tres de las cuales fueron particularmente importantes para la codificación de la rama de actividad económica: la realización de un relevamiento digital autogestionado optativo seguido por el barrido territorial, el empleo de un clasificador adaptado a los establecimientos económicos en lugar de a hogares y la eliminación de algunas preguntas del bloque económico, particularmente la de ocupación.

Dadas las mejoras introducidas en el SiCI, el porcentaje de codificación automática aumentó de manera significativa. Esto, a su vez, permitió disminuir el número de personal de codificación necesario (se habían previsto 150 personas y se necesitaron solo 80 personas) sin aumentar los tiempos de codificación, lo que implicó un ahorro económico y de recursos para el INDEC. Nada de todo esto hubiera sido posible sin el compromiso del personal involucrado, tanto del Instituto como del contratado.

En estas conclusiones, en primer lugar, se debe evaluar el impacto de las tres innovaciones mencionadas.

a) La modalidad dual del relevamiento censal

En general, el relevamiento del censo digital resultó adecuado en cuanto a la calidad de la información de la rama de actividad económica. Aunque la ausencia de la guía de un censista podría implicar mayores dificultades en las respuestas, la información obtenida resultó satisfactoria. Por el contrario, en ocasiones, el relevamiento en campo presentó inconvenientes que deberán ser evaluados en futuros operativos.

b) El empleo del ClaNAE en lugar de la CAES

Previo a la realización del Censo, la CCN sugirió que el ClaNAE no resultaría un clasificador adecuado para el relevamiento de los hogares. Sin embargo, su aplicación fue satisfactoria debido a la labor de la CCN en el trabajo de análisis y adecuación del clasificador para el operativo⁶³. Se codificó aproximadamente la mitad de los casos a la apertura máxima de 5 dígitos, aunque hubo que resignar una codificación homogénea en cuanto a la desagregación. La distribución de la codificación fue: a 5 dígitos, 53,3%; a 4 dígitos, 8,2%; a 3 dígitos, 1,7%; a 2 dígitos, 14,7%; y a nivel de sección, 22,1%.

c) La eliminación de preguntas del bloque económico

En base a la experiencia adquirida, para un próximo operativo censal, se debe volver a evaluar la selección de preguntas del bloque económico.

⁶³ Ver apartado 4.2.1

Además de estas innovaciones principales incorporadas al Censo, otros temas merecen ser evaluados:

d) El lugar donde se desarrolló la codificación

Aunque siempre se habían realizado los procesos de codificación censal en algún sitio relativamente cercano al edificio central del INDEC, dentro de la Ciudad Autónoma de Buenos Aires, la elección del campus de la Universidad de La Matanza tuvo sus pros y sus contras. Por un lado, sumó tiempo adicional de desplazamiento del equipo técnico de la CCN, que iba diariamente y en dos turnos, lo que obligó a diagramar un plan de acción coordinado con el área de logística del INDEC. Por otro, el espacio físico provisto por la Universidad y sus instalaciones resultó adecuado para llevar adelante la tarea.

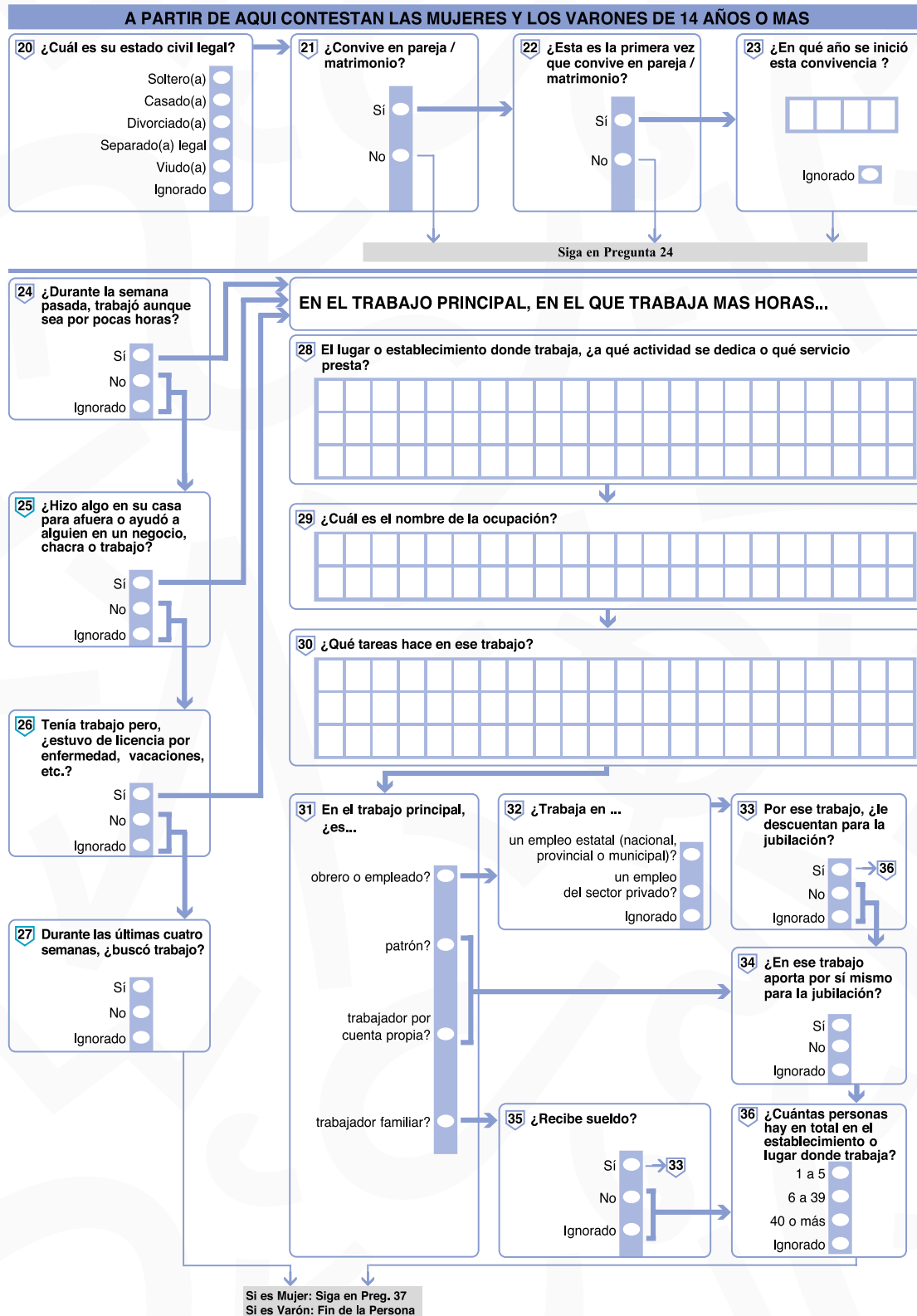
e) La lectura de las cédulas censales

Desde el Censo Nacional de Población, Hogares y Viviendas 2001, las cédulas censales y las preguntas abiertas se escanean con lectoras ópticas que tienen reconocimiento de caracteres. Resulta fundamental contar con una buena lectura para no distorsionar las declaraciones de las personas censadas. En esta oportunidad, el escaneo resultó deficiente, con las subsecuentes demoras y con un porcentaje de información no codificable, como se adelantó.

f) La capacitación del equipo de censistas

Un tema para considerar en cualquier relevamiento a hogares es la formación de quienes lo llevan a cabo en terreno. En este sentido, hubo algunos inconvenientes ya que el tiempo que se invierte para capacitar suele ser escaso en los censos de población. Este tema se hace más notorio en las preguntas abiertas. Además, en esta ocasión, se modificó la metodología de capacitación: históricamente, la dictaban docentes, y, esta vez, se trató de una capacitación virtual autoasistida que, aunque abreviaba el período de capacitación, podría haber influido en el desempeño del personal encargado de censar.

Anexo 3. Diseño del formulario del bloque económico en el Censo 2001



Anexo 4. Proceso de codificación de la rama de actividad en el Censo 2022

El siguiente diagrama sintetiza el proceso de codificación de la rama de actividad económica (pregunta 34). La CCN esbozó un diseño de cómo debía ser dicho proceso, que fue ajustado y simplificado de acuerdo con lo que sucedió realmente.

Diagrama del proceso de codificación de rama de actividad en el Censo 2022



Nota: el color azul corresponde a datos; el verde, a control de calidad; y el marrón, a acciones y procesos.

Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

Anexo 5. Esquema de la secuencia general de pasos en la codificación de la rama de actividad

El siguiente es un esquema de lo desarrollado en el punto 4.2.2 de este informe.

Codificación de la pregunta 34	
paso 1	Si p30 = 1 → CODACT = 97000
paso 2	Se ejecuta el SiCI → CODACT = códigos asignados automáticamente
paso 3	Codificación asistida → CODACT = códigos asignados por los codificadores
paso 4	Si CODACT = V o Z y si p33 = 3 → CODACT = G
	Si CODACT = V o Z y si p33 = 4 → CODACT = C
	Si CODACT = V o Z y si p33 = 6 → CODACT = F
	Si CODACT = V o Z y si p33 = 7 → CODACT = H
	Si CODACT = V o Z y si p33 = 8 → CODACT = I
	Si CODACT = V o Z y si p33 = 9 → CODACT = K

Tratamiento de la pregunta 33	
paso 1	Conversión de CODACT a letra ClaNAE
paso 2	Conversión de letra ClaNAE a p33corregida

Consistencias y corrección de las preguntas cerradas del bloque económico (Trabajo realizado por DNESP)	
paso 1	Corrección de Condición de actividad (quitar de ocupados si CODACT=W)
paso 2	Corrección de Categoría ocupacional (si p30<>1 y CODACT=97000, entonces p30=1)

Anexo 6. Tabla de conversión y correspondencia de los valores consignados en la pregunta 34 (CODACT) a letra ClaNAE

CODACT	Sección ClaNAE
Si CODACT="A"	A
Si CODACT>="01".and.CODACT<"05"	A
Si CODACT="B"	B
Si CODACT>="05".and.CODACT<"10"	B
Si CODACT="C"	C
Si CODACT>="10".and.CODACT<"35"	C
Si CODACT="D"	D
Si CODACT>="35".and.CODACT<"36"	D
Si CODACT="E"	E
Si CODACT>="36".and.CODACT<"41"	E
Si CODACT="F"	F
Si CODACT>="41".and.CODACT<"45"	F
Si CODACT="G"	G
Si CODACT>="45".and.CODACT<"49"	G
Si CODACT="H"	H
Si CODACT>="49".and.CODACT<"55"	H
Si CODACT="I"	I
Si CODACT>="55".and.CODACT<"58"	I
Si CODACT="J"	J
Si CODACT>="58".and.CODACT<"64"	J
Si CODACT="K"	K
Si CODACT>="64".and.CODACT<"68"	K

CODACT	Sección ClaNAE
Si CODACT="L"	L
Si CODACT>="68".and.CODACT<"69"	L
Si CODACT="M"	M
Si CODACT>="69".and.CODACT<"77"	M
Si CODACT="N"	N
Si CODACT>="77".and.CODACT<"84"	N
Si CODACT="O"	O
Si CODACT>="84".and.CODACT<"85"	O
Si CODACT="P"	P
Si CODACT>="85".and.CODACT<"86"	P
Si CODACT="Q"	Q
Si CODACT>="86".and.CODACT<"90"	Q
Si CODACT="R"	R
Si CODACT>="90".and.CODACT<"94"	R
Si CODACT="S"	S
Si CODACT>="94".and.CODACT<"97"	S
Si CODACT="T"	T
Si CODACT>="97".and.CODACT<"99"	T
Si CODACT="U"	U
Si CODACT="99"	U
Si CODACT="V"	V
Si CODACT="Z"	Z
Si CODACT="W"	W

Anexo 7. Tabla de conversión y correspondencia de sección ClaNAE a la pregunta 33 corregida

Sección ClaNAE		p33corregida ⁽¹⁾
A		5
B		5
C		4
D		10
E		10
F		6
G		3
H		7
I		8
J		11
K		9
L		11
M		11
N		11
O		2
P	Si p33=1	1
P	Si p33=2	2
P	Si p33 distinto de 1 y 2	12*
Q	Si p33=1	1
Q	Si p33=2	2
Q	Si p33 distinto de 1 y 2	12*
R		11
S		11
T		11
U		11
V ⁽²⁾		p33original
Z ⁽²⁾		p33original
W ⁽²⁾		(Se excluye de los ocupados)

⁽¹⁾ p33corregida = 12: educación o salud sin determinar si es pública o privada.

⁽²⁾ El ClaNAE no recoge estas secciones porque no expresan ramas de actividad.

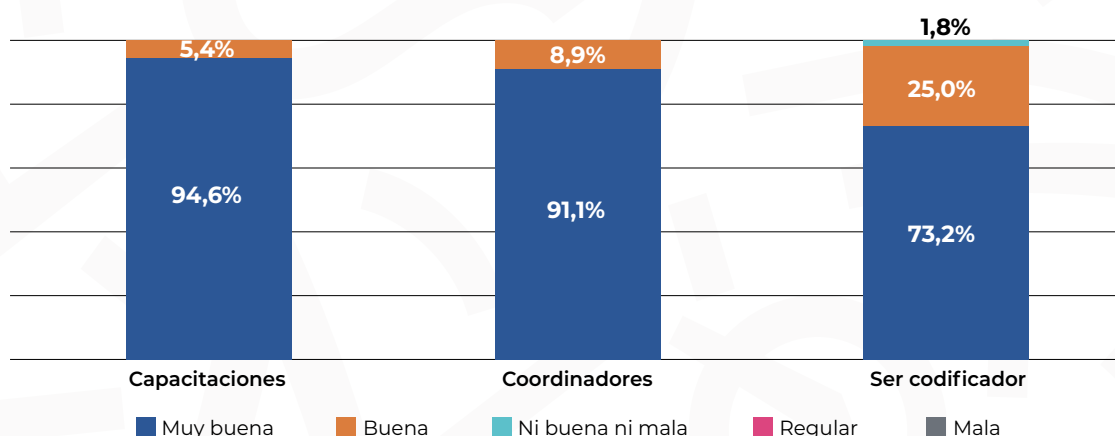
Anexo 8. Evaluación del proceso de codificación por parte de las codificadoras y los codificadores

Tras la primera instrucción, se realizó una encuesta anónima y obligatoria para optimizar la preparación y el dictado de próximas instancias de formación. Al terminar con el operativo, se realizó una segunda encuesta para indagar sobre el proceso en términos generales.

A continuación, se presentan algunos datos que ilustran las perspectivas de quienes codificaron respecto al operativo en general y a las capacitaciones.

En el gráfico 1, se despliega información sobre la valoración de las y los codificadoras en cuanto a los aspectos generales que componen la labor, que remiten a las capacitaciones de la tarea que realizaron, al rol del equipo de la CCN como personal a cargo de asesorar y guiar el trabajo cotidianamente y al rol propio en lo que refiere al desarrollo técnico de su labor de analizar y codificar la información.

Gráfico 1. Valoración de las personas codificadoras de rama de actividad del Censo 2022 sobre aspectos generales de la labor de codificación



Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

El gráfico 1 muestra que las capacitaciones y el rol del equipo de coordinación fueron los aspectos mejor valorados.

El 94,6% le otorgó la mejor valoración a las capacitaciones, seguido por el 5,4% que otorgó la valoración “buena”. Al profundizar en los motivos, un testimonio afirmó: “la capacitación inicial y el acompañamiento constante de los coordinadores generó un buen ambiente y la posibilidad de reducir significativamente el margen de error en los casos a evaluar”. Otra persona encuestada dijo: “me parecieron muy buenas y necesarias las capacitaciones colectivas para mejorar la calidad de codificación” y en línea con esta valoración, otra respuesta fue: “en cada momento que se realizó una recapacitación me permitió mejorar aún más mi trabajo”.

En cuanto al rol del equipo de coordinación, el porcentaje de respuesta positiva fue alto (91,1%, muy buena y 8,9%, buena). Una de las respuestas que sintetiza esta tendencia dice: “excelente. Además, de principio a fin durante las jornadas laborales fuimos aprendiendo diariamente del constante acompañamiento de todos los coordinadores”. Otro testimonio menciona: “los coordinadores siempre acompañaron, no solo en lo que es el trabajo, sino que también en otros aspectos más personales, fueron muy comprensibles”.

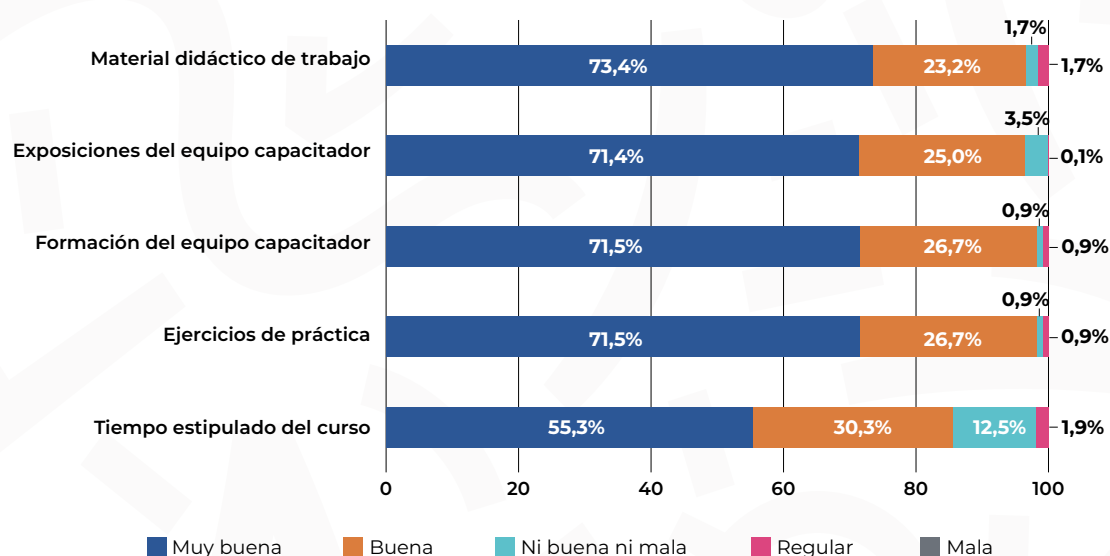
Se observa así la relevancia que se le otorgó a trabajar en un entorno ameno, que estimula la mejora de la calidad: “considero que la capacitación y el acompañamiento constante de coordinadores generó un buen ambiente y la posibilidad de reducir significativamente el margen de error en los casos a evaluar”.

La valoración de la tarea de codificar tuvo porcentajes de respuestas más variadas. El 73,2% expresó que le pareció una experiencia “muy buena”, mientras que el 25,0% que la caracterizó como “buena” y 1,8% como “ni buena ni mala”.

Por un lado, los motivos que se relacionan con la mejor valoración son argumentos similares a los que explicaban la buena mirada sobre las capacitaciones y el equipo coordinador, es decir, la identificación del acompañamiento de la coordinación y de las capacitaciones como mecanismos para alcanzar un ambiente de trabajo óptimo: “aunque desconocía las especificidades del trabajo, el apoyo de los supervisores, su capacitación y colaboración han traído aparejado que pueda desarrollar mi potencial laboral y el trabajo en equipo”. Por otro lado, en el caso en que la respuesta fue “ni buena ni mala” no se brindó una justificación para dar cuenta del motivo de dicha valoración.

El gráfico 2 presenta la evaluación realizada sobre las estrategias de capacitación.

Gráfico 2. Evaluación de las personas codificadoras de rama de actividad del Censo 2022 del aporte de las estrategias de capacitación



Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

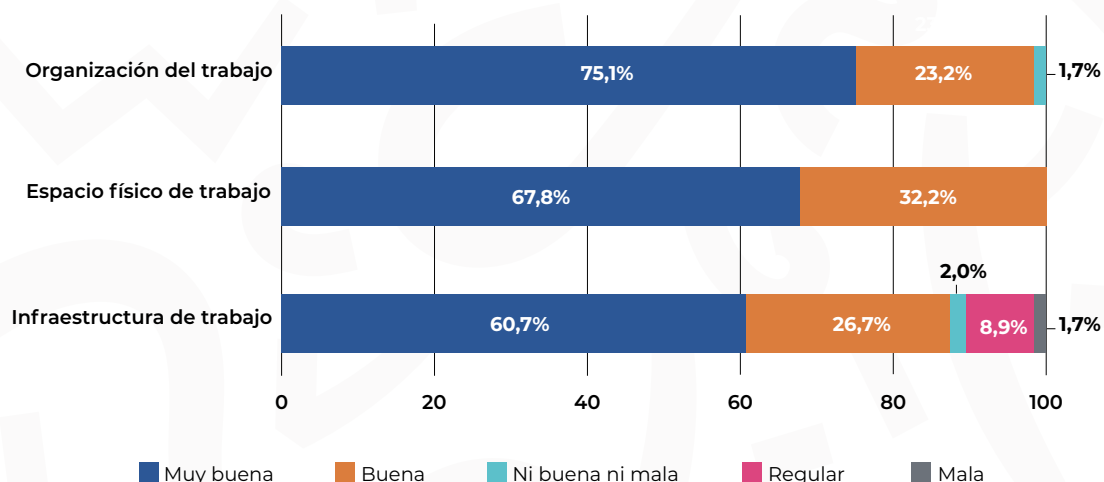
El gráfico 2 muestra que las estrategias de capacitación con mejor valoración fueron aquellas vinculadas con el material didáctico de trabajo: “la estructura de la ClaNAE muy buena, muy completa, fácil de interpretar, al igual que las notas explicativas. El árbol muy útil al principio”. Los materiales de trabajo fueron pensados y diseñados tomando en consideración la extensión del clasificador. Así, se buscó generar un abordaje didáctico para brindar cercanía en la iniciación del trabajo de codificación.

El equipo capacitador realizó sus exposiciones y planteó los ejercicios de práctica para la codificación utilizando como soporte estos materiales. Estos aspectos, al igual que la formación del equipo técnico, se evaluaron de forma positiva en general: “considero que fue una muy buena experiencia laboral, se nos otorgaron las herramientas necesarias para dicho trabajo. Los coordinadores siempre atentos a nuestras consultas, aprendizaje y en busca de una mejora continua de las herramientas, sabiendo adaptarse a las necesidades que iban surgiendo en cada caso a codificar”.

Respecto al tiempo destinado para el desarrollo de las capacitaciones, este aspecto es el que tuvo la mayor dispersión en cuanto a la valoración. Por un lado, el 55,3% le atribuyó la mejor valoración, seguido del 30,3% que lo encontraron muy adecuado. Por otro lado, el 12,5% le otorgó una puntuación media y el 1,9% una valoración regular. No se encontró información en la encuesta sobre las razones de estas respuestas. Sin embargo, a lo largo del operativo, se recibieron solicitudes de extensión de la duración de las capacitaciones. Debido a los tiempos disponibles y comprometidos para todo el proceso, no se podían prolongar más que lo estipulado. Además, desde la coordinación se consideró que alargarlas daría lugar a saturación en el aprendizaje.

Por último, se indagó acerca de aspectos relacionados a las condiciones generales del espacio laboral. Esta información se presenta en el gráfico 3.

Gráfico 3. Calificación de las personas codificadoras de rama de actividad del Censo 2022 acerca de aspectos físicos, materiales y organizacionales del espacio de trabajo



Fuente: INDEC, Dirección Nacional de Metodología e Infraestructura Estadística. Coordinación de Clasificadores y Nomenclaturas.

La organización del trabajo remite a la propuesta de manejo del tiempo de la CCN (horarios a cumplir, un recreo por turno, etcétera). El 75,1% de las personas encuestadas manifestaron estar muy de acuerdo con la propuesta, frente a 23,2% que se mostró conforme y 1,7% que no la encontró ni buena ni mala.

En lo que respecta al espacio físico de trabajo específico en el aula, el uso de mobiliario y de las computadoras, el 67,8% lo calificó como “muy bueno” y el 32,2% restante como “bueno”.

Por último, sobre la infraestructura de trabajo, el campus de la Universidad y sus instalaciones (baño, comedor, etcétera), el 87,4% le atribuyó las valoraciones más positivas: el 60,7% la calificó como “muy buena” y 26,7% como “buena”. El 2% la encontró “ni muy buena ni muy mala” y el 10,6% restante la consideró “mala” (8,9%) y “muy mala” (1,7%).

Uno de los testimonios que ilustra algunos de estos aspectos fue el siguiente: “el espacio físico fue muy cómodo, a pesar de algunas cosas como la limpieza del espacio. Con respecto a la infraestructura del lugar, en el comedor se dificultaba encontrar un lugar libre y se perdía bastante tiempo”.

5. Equipos de trabajo para la codificación

Dirección Nacional de Metodología e infraestructura estadística

Gerardo Mitas

Coordinación de Clasificadores y Nomenclaturas

Claudio La Rocca

Coordinación del trabajo de codificación en UNLaM: Patricia Höxter y Claudio La Rocca

Equipo docente de capacitación: Anastasia Cenóz, Yanina Conca, Sergio De Marco, Cristina Dellegrazie, Javier Domench, Dolores González, Leandro Rodríguez, Marcelo Rosenzveig, Norma Vincenti

Equipo de coordinación y supervisión: Laura Acevedo, Anastasia Cenóz, Yanina Conca, Sergio De Marco, Cristina Dellegrazie, Javier Domench, Marcelo Esses, Sonia Fernández Gallo, Diego Galante, Dolores González, Leandro Rodríguez, Marcelo Rosenzveig, Roberto Ruiz y Norma Vincenti

Procesamiento informático

Gerardo Graniel

Equipo de codificación de Pueblos indígenas u originarios⁶⁴

Alejo Azar, Cacique del pueblo Diaguita del Valle de Tafí; Ángela Marcela Jaramillo, del pueblo Kolla; Ángela Nahuelquir, del pueblo Mapuche; Bárbara Manasse, del Consejo de Ancianos pueblo Diaguita del Mollar; Carlos Aníbal Loza, del pueblo Comechingón; Carlos Reinaldo Villegas, del pueblo Huarpe, del Consejo Joven Comunidad Huarpe Pinkanta; David Torres, del pueblo Lule; Franco Emanuel Gil, del pueblo Huarpe, representante por San Juan en el CPI; Gabriel Jofre, del pueblo Mapuche; Héctor Santomil, del pueblo Charrúa; Julia Martha Ceballos, autoridad del pueblo Comechingón; Luis Alberto Romero, Werken del pueblo Mapuche; Luis Pilquiman, del pueblo Mapuche y vicepresidente del INAI; María Eugenia Talavera, del pueblo Fiscara; Miguel Roque Gil, del pueblo Huarpe; N. G. Daniel Huircapan, Gama del pueblo Günün a Kūna; Roxana Montenegro, del pueblo Huarpe; Silvia Beatriz González, delegada CEAPI del pueblo Guaraní; y Silvina Eleonora Rivas, Inanlonco del pueblo Mapuche

Equipo de codificación de Rama de actividad económica

Abigail Ralli, Agustín Pérez Tedeschi, Agustina Nobile, Alberto García, Aldana Redin Posse, Analía Vidal, Andrea Varas, Andrés Migliano, Axel Sundberg, Belén Brandolini, Camila Belen Texidor, Camila Lorenzon, Camila Rivas, Candela Aylén Civalero, Catalina Martínez Torres, Cecilia Farias Viale, Cecilia Jourdan, Cesar König, Chiara Gandine Re Delle, Daniela Guzmán, Débora Elizabeth Neyra, Eladia Rodríguez, Elías Correa, Elizabeth Laura Coronado, Emmanuel Cueto, Estefanía Martone, Eugenia Aguilar, Ezequiel Ojeda Lopaczenok, Facundo Kormann, Facundo Tassarolo, Fiamma Pedone, Fiorella Páez, Florencia Ariza, Florencia Dalesio, Franco Velastiqui, Gonzalo Daniel Cueto, Guillermo Coronel, Hilda Mamani, Ignacio Barborini, Jessica Simoncelli, Joaquín Noe, Johana Natalia Jaimez, Johanna Falgione, Jorge Sánchez, Josué Alvaríño Keiserman, Juan Alberto Cabelluzzi, Juan Cruz Antigués, Juan Guerrini, Julieta Millares, Leonardo Calbuyahue, Leonardo Jara, Leonel Gabriel Iglesias, Lihuen Ocampos, Lucas Galli, Lucia Nani, Luciano Barabini, Magali Verón, Malena Ayala, Malena Forte, Malena Jara, Mara Acorroni, Marcela Moleres, María Belén Costilla, María Sol Groba, Mateo Cordon, Matías Algarañaz, Matías Corvalan, Melina Fernández, Micaela Llano, Micaela Maggolino, Milena Benítez, Nicolas Ledesma, Noelia Fernández Sotelo, Ornella Saitta, Paula Casas, Rocío Agüero, Rocío Belen Nemeth, Rodrigo Sebastian Del Vecchio, Romina Romanelli, Ruth Soler, Sergio Latorre, Sofia Aylén González, Sofia Coniglio, Stefano Ramis, Stella Maris Staffa, Tomas Lemaire, Tomy Alanis, Trinidad Serrano, Valentín Subbioto, Victoria Carralbal y Yamila Peñaloza

⁶⁴ Equipo de personas expertas y referentes de PPIL propuesto por el INAI.

6. Bibliografía

Instituto Nacional de Asuntos Indígenas (INAI) (5 de agosto de 2014). Anexo I. En *Resolución 737/2014*. Ministerio de Desarrollo Social.

_____. (2021). *Respuesta a la Codificación de Pueblos solicitada por INDEC*. NO-2021-68423575-APN-INAI#MJ.

Instituto Nacional de Estadística y Censos (INDEC) (2005). Capítulo 1. Antecedentes. En *Encuesta Complementaria de Pueblos Indígenas (ECPI) 2004 - 2005*. https://www.indec.gob.ar/micro_sitios/webcenso/ECPI/ECPI%20-%20Antecedentes.pdf

_____. (2005). Presentación. *Encuesta Complementaria de Pueblos Indígenas (ECPI)*. https://www.indec.gob.ar/micro_sitios/webcenso/ECPI/index_ecpi.asp

_____. (2019). *Informe de evaluación de la segunda prueba piloto. Hacia la Ronda 2020* [Documento de Trabajo N° 29]. https://www.indec.gob.ar/ftp/cuadros/publicaciones/censo_2020_informe_2PP.pdf

_____. (2020). *Informe sobre la codificación de la pregunta sobre pueblos originarios de la Segunda prueba piloto del Censo 2020*.

_____. (2023). *Manual de capacitación de Pueblos indígenas*.

_____. (s.f.). *Preparación del censo experimental - Hacia la ronda 2020. Diseño conceptual del censo 2020. Listado de pueblos indígenas* [Documento de Trabajo s.n.].

Siglas y acrónimos

II PP	Segunda Prueba Piloto
CAES	Clasificación de Actividades Económicas para Encuestas Sociodemográficas
CCN	Coordinación de Clasificadores y Nomenclaturas
CE	Censo Experimental
CIU	Clasificación Industrial Internacional Uniforme
ClaNAE	Clasificación Nacional de Actividades Económicas
CNPhyV	Censo Nacional de Población, Hogares y Viviendas
DNMIE	Dirección Nacional de Metodología e Infraestructura Estadística
DECE	Dirección de Estadísticas del Comercio Exterior
DEP	Dirección de Estadísticas Poblacionales
DNESP	Dirección Nacional de Estadísticas Sociales y de Población
ECPI	Encuesta Complementaria de Pueblos Indígenas
INAI	Instituto Nacional de Asuntos Indígenas
INDEC	Instituto Nacional de Estadística y Censos
NPI	Nomenclador de Pueblos Indígenas
PI	Pueblo indígena
PPII	Preguntas sobre pueblos indígenas u originarios
SiCI	Sistema de Codificación Informatizada
OIT	Organización Internacional del Trabajo
UNLaM	Universidad Nacional de La Matanza