

CONVERGENCIA DE LA ADMINISTRACIÓN TRADICIONAL Y LA E-ADMINISTRACIÓN EN LOS CENSOS DE POBLACIÓN Y VIVIENDAS DE 2001

Joaquín Agurruza Mutuberría

(INE)

1. Introducción

Con carácter decenal, el Instituto Nacional de Estadística viene internacionalmente obligado a realizar entre los residentes en España (tanto españoles como extranjeros) los Censos de Población y Viviendas

De acuerdo con la legislación estadística, corresponde hacer dicha operación en los años terminados en uno.

Como no podía ser de otra forma, en los Censos de Población y Viviendas 2001, las nuevas tecnologías han jugado un papel fundamental a lo largo de todas las fases.

En las fases previas a la entrega y posterior recogida de los cuestionarios censales, las herramientas informáticas han permitido diseñar de una forma eficiente los diferentes modelos de cuestionarios, procesar en tiempos muy reducidos los datos de las dos pruebas piloto realizadas (de forma que sus resultados se han podido utilizar para mejorar el diseño definitivo de la operación) y preparar, a partir de los datos padronales, los ficheros que se han utilizado en la personalización de los cuestionarios del censo.

Los tratamientos informáticos han permitido también eliminar la tradicional operación de recorrido del territorio un año antes de los Censos de Población y Viviendas, que servía para formar los Censos de Edificios y Locales y para preparar los cuadernos de recorrido de los agentes censales (actualizaciones del seccionado y del callejero).



En este censo, esa costosa operación se ha sustituido ventajosamente por tratamientos informáticos aplicados a registros administrativos, principalmente padronales y catastrales, como resultado de los cuales se han podido preimprimir tanto los cuadernos de recorrido como la información de datos personales y de domicilios de las personas empadronadas.

La gestión de la contratación del personal eventual necesario para realizar la recogida, agentes censales y encargados de grupo (mas de 40 mil personas), se ha realizado de manera descentralizada en las Delegaciones Provinciales del INE.

Ello ha posibilitando los aspectos materiales de la propia contratación (generar diversas copias -incluida las copias básicas- de más de 40.000 contratos en menos de 5 días); gestión de altas y bajas a la Seguridad Social (a través del sistema RED) y la gestión de los componentes variables de las nóminas de dichos trabajadores.

Durante la propia operación de recogida censal, las nuevas tecnologías han estado igualmente presentes. Así, las 500 oficinas comarcales en las que se ha descentralizado la operación de la recogida, han estado comunicadas con las Delegaciones Provinciales y con los Servicios Centrales del INE por una red privada (Red Privada de Censos), basada en telefonía móvil.

Cada una de las oficinas comarcales se informatizó con dos ordenadores personales, uno de ellos con comunicación a la red privada de censos, una impresora, y un lector de código de barras. Esta infraestructura ha resultado fundamental para facilitar el control de la operación de la recogida, para poder enviar rápidamente instrucciones a todas las oficinas censales, y para facilitar la información de los cupos de trabajo realizados por cada agente para calcular la parte variable del pago de sus nóminas.

A pesar de no haber precedentes en los que se permitiese la cumplimentación, con carácter cuasi-universal, de los cuestionarios por internet, el INE decidió asumir dicho reto.

Así se hizo, estableciendo un procedimiento que compatibilizara las mayores medidas de seguridad en cuanto al acceso a la información de los datos de carácter personal, con la cumplimentación del censo por internet a todas aquellas personas previamente empadronadas donde residían a la fecha de referencia utilizada para hacer la preimpresión de la información de datos padronales.

Además, con el objetivo de universalidad en mente se diseñaron las páginas Web **permitiendo la cumplimentación por personas con discapacidades visuales**. Finalmente el número de hogares que han cumplimentado el censo por internet ha sido de **13.818**, que representa alrededor de un uno por mil del total de los hogares existentes. En el apartado 2 de este documento se hace un breve resumen de las características técnicas de los sistemas utilizados.



Para coordinar la recogida de la información a la manera tradicional con la recogida a través de Internet ha sido crucial la interconexión de los sistemas informáticos de las oficinas comarcales con las Delegaciones Provinciales y los Servicios Centrales a través de la Red Privada.

Esta interconexión ha permitido conocer qué hogares habían realizado la cumplimentación por internet, **evitando de esta forma tener que recoger el cuestionario en papel**. Adicionalmente, han permitido controlar los envíos de los cuestionarios cumplimentados en papel al centro de producción censal constituido para realizar la captura y procesamiento informático de dicha documentación, facilitando que se iniciasen los trabajos en soporte papel sin necesidad de esperar a que se diese por cerrada la recogida en cada una de las secciones censales.

Para la obtención de la información de los cuestionarios en papel, ha sido preciso establecer un Centro de producción del INE, expresamente dedicado a realizar este trabajo, realizándose las mejoras de las instalaciones del edificio que lo alberga y el montaje de la arquitectura de los sistemas y de las aplicaciones necesarias.

Se finaliza con una breve descripción del resto de los tratamientos informáticos a que se someten los datos recogidos, que permitirán poner a disposición de los usuarios la información censal.

2. Captura de cuestionarios por internet

Las ideas claves que definieron este proyecto fueron:

- **España ha sido el primer país del mundo que ha permitido la cumplimentación del Censo por Internet de manera general**, entendiendo por tal a todas las personas previamente empadronadas en la vivienda donde residían a la fecha de la preimpresión de los cuestionarios
- La cumplimentación por Internet del Censo se hizo asegurando su confidencialidad y de forma que resultara fácil de realizar; en función de procedimientos de seguridad acordes con la información a cumplimentar
- Se primó la cumplimentación por Internet del Censo mediante la entrega gratuita de datos estadísticos sobre la distribución geográfica de un apellido que el informante solicite (salvaguardando la confidencialidad estadística)



- El INE posibilitó la cumplimentación del Censo de Población por Internet a personas con discapacidades visuales o de otro tipo con dificultades para el manejo del ordenador

A continuación se hace una breve descripción de la forma en que se abordó este proceso:

Según el diseño de la operación censal española, en los cuestionarios en papel que se distribuyeron a los hogares españoles, figuró preimpresa determinada información individual de cada ciudadano, obtenida de la base de datos padronales.

Se estableció el cuestionario censal en un servidor web seguro SSL 3 (con la dirección (<https://www.censos2001.es>), con un mecanismo de firma electrónica avanzada (certificados X.509 de clase 2, mediante convenio con la FNMT-RCM), para aquellos usuarios que requiriesen modificar sus datos padronales, complementada con un procedimiento de autenticación alternativo en caso contrario, para proceder a la cumplimentación del censo propiamente dicho.

El mecanismo alternativo de autenticación se basó en los siguientes identificadores:

- 1) CLAVE1: Código identificativo incluido en cada sobre conteniendo los cuestionarios censales;
- 2) CLAVE2: Contraseña asociada a la cumplimentación por Internet, también incluida en cada sobre censal;
- 3) El DNI de una de las personas incluidas en la hoja de datos padronales (DNI también preimpreso en el cuestionario) y
- 4) El nombre del padre y de la madre tal y como figuran en el DNI de determinada persona del hogar (esta información no figuraba preimpresa en el cuestionario censal).

CLAVE1 Y CLAVE2 fueron distintas para cada vivienda.

Relacionado directamente con la seguridad, se implementaron un conjunto de medidas orientadas a realizar el control de accesos incorrectos, de intentos de fraude, bloqueos y desbloqueos de cuestionarios, etc.

El servidor web ofreció la posibilidad de cumplimentar los cuestionarios en los distintos idiomas cooficiales en España y en algunos otros extranjeros (francés, inglés, alemán y árabe).

Se incluyeron un conjunto de normas de edición del cuestionario en formato web; es decir el conjunto de edits necesario para asegurar la calidad y la consistencia de cada uno de los cuestionarios cumplimentados por Internet, comu-



nicando al usuario cualquier problema que impida la aceptación final de dicha cumplimentación, para que aquel procediera a corregir los errores correspondientes.

Asimismo, el sistema permitió que el usuario pudiera interrumpir su trabajo de cumplimentación del cuestionario, pudiéndolo continuar posteriormente. En el momento en que la cumplimentación se hubiera finalizado de forma totalmente correcta, el sistema proporcionaba al usuario un número que servía de recibo o de comprobante de que dicha cumplimentación completa se había producido.

Se implementaron los mecanismos necesarios para la comunicación a establecer con las Oficinas Comarcales y con las Delegaciones Provinciales del INE, para la actualización (a partir de los datos identificativos de los cuestionarios recogidos por Internet) de sus respectivas bases locales conteniendo los datos identificativos de los cuestionarios recogidos en papel y de modo que ningún agente censal reclamara los cuestionarios previamente cumplimentados por Internet.

Esa comunicación con las Oficinas Comarcales consideró diferentes posibilidades: como mecanismo básico existió un procedimiento de envío o de descarga, de modo que cada Oficina Comarcal y cada Delegación Provincial dispusiera semanalmente de un fichero conteniendo los datos identificativos de los cuestionarios recogidos por Internet, y alternativamente, procedimientos de consultas por rangos de valores.

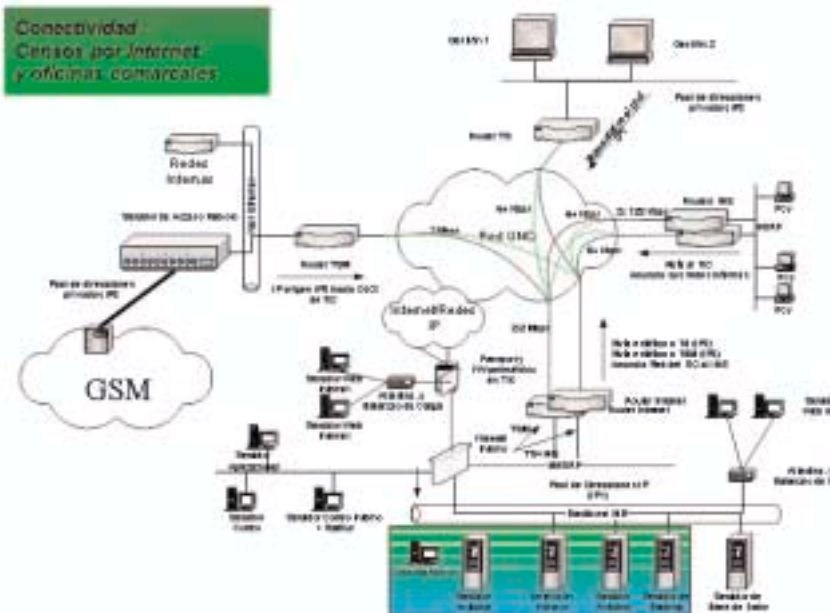
En cuanto a las características de la solución utilizada, fueron las siguientes:

- Caudal de la conexión con internet: CVP de FR Red Uno, con un caudal garantizado de 34 Mbps con backup equivalente.
- Calidad del servicio: Retardo máximo de 120 ms desde la sede del Webhosting al nudo neutro de Espanix
- Clientes: Navegadores con soporte HTML 4.01 y Hojas de estilo en cascada (Internet Explorer 5.x, Opera,) y HTML dinámico, mediante JavaScripts 1.2 y modelo DOM del W3C.
- Solución software basada en una arquitectura de tres capas (Presentación / Aplicación / Datos):
- Tecnología de componentes Enterprise Java Beans
- Servidor de aplicaciones Bea Web Logic,
- Servidor de correo iPlanet Messaging Server



- Servidor de BD Oracle 8i Enterprise Server con Oracle Parallel Server
- Hardware robusto y de alta escalabilidad:
- Servidores SUN E420R clusterizados, con biprocesadores a 450 MHz y a 1 y 2 GB de RAM para los servicios de correo y de aplicaciones
- Identica configuración para el SGBDR con sistemas de almacenamiento clusterizados SUN D1000 de 144 GB
- Firewall perimetral clusterizado basado en servidores SUN 220 450 MHZ y 1 GB RAM y SW de CheckPoint
- Firewall interno clusterizado basado en servidores SUN 220 450 MHZ y 1 GB RAM y SW de Axent Raptor
- Balanceadores de carga ACEdirector 3, de AlteonWebSystems - Nortel

El siguiente gráfico muestra la arquitectura de los sistemas y de las comunicaciones global, del servicio de webhosting:



Arquitectura Webhosting

Detalle Componentes Hardware

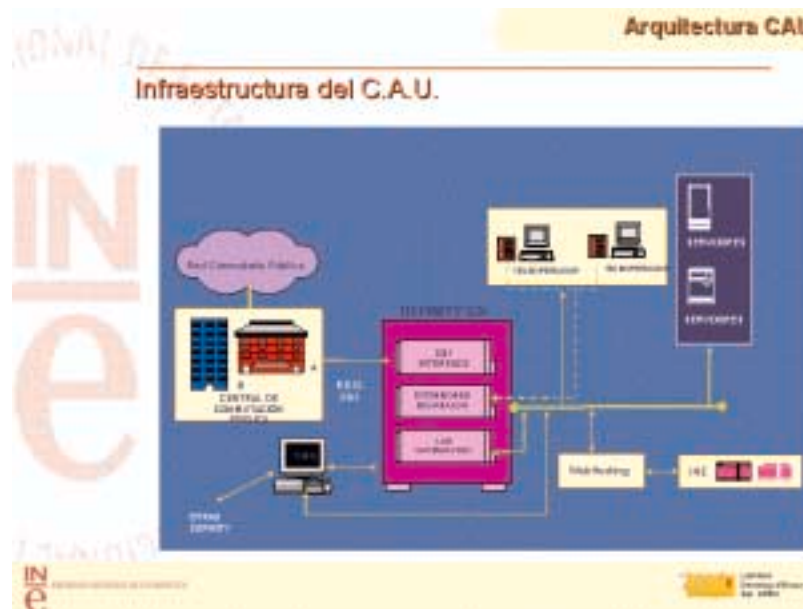
• Arquitectura HW redundante

Servidores web internet	450 Mhz	1 GB RAM	18 GB UltraSCSI
Servidores aplicaciones	450 Mhz	2 GB RAM	18 GB UltraSCSI
Servidores correo	450 Mhz	1 GB RAM	36 GB
Servidores BD	450 Mhz	2 GB RAM	18 GB UltraSCSI
Servidores intranet	450 Mhz	1 GB RAM	18 GB UltraSCSI
Servidores detección intrusos	400 Mhz	512 MB	20 GB EIDE
Servidores antivirus	400 Mhz	512 MB	20 GB EIDE
Servidores Consola administración	400 Mhz	256 MB	20 GB EIDE
Servidores back-up	400 Mhz	256 MB	20 GB EIDE
Servidores consola administración cortafuegos y gestión	400 Mhz	512 MB	36 GB
Cortafuegos	450 Mhz	1 GB	36 GB



Se ha contado con un CAU (Centro de atención al usuario) para recoger todas las llamadas con dudas de los ciudadanos con un caudal garantizado conexión webhosting y CAU: de 128CVP 64Kbps:

- 4 plataforma (Madrid, Coruña, Bilbao, Barcelona)
- Nº medio de teleoperadores: 19 (desde un mínimo diario de 7 a un máximo de 39)
 - Sistema Call Center Multichannel (integración llamadas telefónicas y por Internet – chat, Voz IP e incluso videoconferencia);
 - Call me back
 - Atención formularios web
 - VRU (unidad de respuesta vocal) Desbordamiento y servicio fuera de horario





3. Tratamientos posteriores al Centro de Producción Censal

Cuando acaba el trabajo del centro de producción censal, los cuestionarios censales han sido escaneados, reconocidos y validados. Estos procesos se han apoyado en diccionarios que han permitido a su vez realizar la codificación de aquellas preguntas que admiten una respuesta literal (provincia y municipio de nacimiento o residencia en 1991, actividad, ocupación, etc.)

Sin embargo no se consigue el 100 por 100 de las codificaciones. Tampoco las validaciones que van asociadas a controles de coherencia son necesariamente exhaustivas pues se centran en la eliminación de los errores más importantes. Por este motivo hay que aplicar tratamientos adicionales que permitan la obtención de los ficheros finales del censo plenamente explotables estadísticamente.

Esta fase se realiza de dos maneras alternativas: para los registros correspondientes a comunidades autónomas que han firmado un convenio de colaboración que lo prevé, son los institutos de estadística de estas comunidades los que se responsabilizan de esta codificación, según sus posibilidades, aunque con criterios coherentes con los que utiliza el INE para el resto del Estado.

Para los registros del resto de las comunidades, es el Instituto Nacional de Estadística el que se encarga de la tarea, utilizando una versión actualizada de los procedimientos ya utilizados con éxito en los Censos de 1991, y que consisten básicamente en una codificación automática por aproximación, apoyada en diccionarios que se mejoran progresivamente.

Finalmente, independientemente del origen de los datos (ya sean de los institutos de estadística autonómicos como los procesados por el INE se someten a un único procedimiento de imputación automática ejecutado por el INE (al igual que los tratamientos de codificación automática, este tratamiento se realiza de forma centralizada en la Subdirección de Informática), que tiene por objeto eliminar las inconsistencias y que consta de un proceso de imputación probabilística que mantiene al máximo la información original.

Este procedimiento produce así, el fichero final que se utilizará en las explotaciones estadísticas que realicen tanto el INE como los institutos autonómicos. Con esto se logra una notable economía de medios (al aplicarse un tratamiento único para todos los datos de España), y lo que es igualmente importante, un único fichero final que evite que una misma fuente estadística aporte distintas cuantificaciones de un mismo fenómeno.

Respecto a Censos anteriores, el grado de utilización de la imputación automática va a ser mucho menor, puesto que las depuraciones y controles aplicadas en el centro de producción hacen que la calidad de los datos que llegue a ella



sea mucho mejor. El INE volverá a realizar la imputación automática utilizando el sistema DIA, desarrollado por el INE y ya aplicado en 1991 y en otras encuestas como la EPA.

Se tiene previsto realizar un sistema de difusión de los resultados obtenidos de manera complementaria por tres procedimientos atendiendo a los niveles de desagregación geográfica. Así, para los resultados a nivel provincial o superior la difusión se hará impresa y por internet, a nivel municipal en CD ROM e internet y para el nivel inframunicipal exclusivamente por internet.

La información a publicar se extraerá de la Base de Datos de microdatos mediante técnicas DataWarehouse, que necesariamente deberán ser particularizadas para su utilización en un entorno estadístico para que cumplan los criterios de seguridad específicos (secreto estadístico,...).

Por último, reseñar que se hará una evaluación de la calidad del Censo mediante un contraste de las informaciones del propio Censo con las correspondientes de la EPA, que permitirá medir los errores de cobertura y de contenido de esta operación.