

10. Procesamiento informático

Informatización de los datos censales

En este apartado se va a producir una de las principales novedades tecnológicas de los próximos Censos: la **utilización de escáneres de alto rendimiento¹ para capturar previamente la imagen de los cuestionarios y, a continuación, aplicar técnicas de reconocimiento óptico**; y no sólo de marcas (como en el Censo de Edificios de 1990), sino también de números y letras manuscritas.

En concreto, el método será de tipo mixto: para los literales manuscritos (en letras de imprenta), donde la efectividad del reconocimiento es menor, se realizará una comparación posterior entre los *valores reconocidos* y su imagen real; para mayor eficiencia, esta comparación será selectiva, comprobando únicamente los literales cuyos valores reconocidos no se encuentren en determinados diccionarios auxiliares (geográficos, de ocupaciones, de actividades.) o cuyo índice de fiabilidad del reconocimiento sea bajo. Y sólo en los casos de discrepancia entre valores reconocidos y reales, será necesario introducir estos vía teclado, con la consiguiente agilización del proceso en comparación con el método tradicional.

La utilización de escáneres permitirá además un ahorro importante en los costes de almacenamiento de la documentación censal, al permitir la destrucción del soporte físico en papel de manera casi inmediata: su imagen será más útil y mucho más fácil de custodiar, cubriendo por esta vía unos de los aspectos más complejos de la protección de los datos censales.

Tratamientos informáticos

Los tratamientos informáticos (procesos de depuración, codificación, tabulación.) que se deben aplicar a los datos censales están fuertemente condicionados por el enorme volumen de información a procesar. Así, procedimientos, sobre todo con cierto componente manual, que funcionan perfectamente en una encuesta por muestreo pueden ser totalmente inadecuados para unos censos.

Por otra parte, los usuarios demandan insistentemente una reducción sustancial del tiempo que deben esperar hasta que los distintos resultados censales ven a la luz². Ambos factores confluyen en que **los tratamientos censales**, aparte de asegurar un nivel de calidad aceptable, **deben ser, ante todo, rápidos**.

El tratamiento informático de los próximos Censos estará también condicionado por la necesidad de aumentar la coordinación entre los Institutos Autonómicos de Estadística y el INE en estas tareas, con el propósito de conseguir un uso más racional de los recursos³, con el consiguiente ahorro de costes y plazos.

Así, el nuevo modelo de colaboración consiste en una división de funciones entre el INE y los institutos autonómicos, de tal manera que cada fase informática sea realizada una, y sólo una, vez. Las autonomías interesadas en participar en este aspecto censal asumirán la supervisión de las tareas de informatización de la documentación censal de sus respectivos territorios, de acuerdo con unas

¹ Los mejores modelos actuales superan ampliamente la barrera de los 100 DIN-A4 por minuto.

² Propósito que va siendo mencionando una y otra vez, cuando se habla de cómo enfocar cada aspecto censal relevante, lo que demuestra bien a las claras que debe ser uno de los ejes principales de la próxima operación censal.

³ Evitando procesamientos en paralelo y, en particular, la corrección descoordinada de los mismos errores, mediante procedimientos, en general, muy similares pero que no terminan, lógicamente, arrojando los mismos resultados. Este modelo, aplicado en 1991, fue un buen punto de partida para empezar a colaborar en esta materia, pero debe ser mejorado sustancialmente con vistas a 2001.

normas generales, previamente consensuadas, y en la que estarán integradas unas primeras tareas de depuración y codificación.

El criterio para decidir qué procesos conviene concentrar en este entorno deberá ser la conveniencia de acudir a las imágenes escaneadas para dar una mejor solución. El objetivo de estos tratamientos será, por tanto, mejorar los resultados del procedimiento de captura (en el sentido de que refleje lo más fielmente posible las respuestas contenidas en los cuestionarios físicos) y sólo excepcionalmente modificar su contenido.

Estos primeros tratamientos serán de enorme importancia, porque, de ser realizados adecuadamente, dejarán la información parcialmente depurada y codificada, y con un grado de calidad suficiente como para permitir que los posteriores tratamientos tengan un componente manual mínimo.

Al acabar esta fase, el INE, responsable último de la fase de captura, enviará a cada instituto autonómico participante el fichero resultante, incluyendo ya la fase de consolidación interna de todos los datos padronales entre sí. A partir de ese fichero será posible difundir avances de resultados, indicando siempre el carácter provisional de las cifras. Algunos institutos asumirán a continuación labores de depuración y codificación intermedias (la principal, codificar los literales que no hayan podido codificarse en el centro de escaneo) y, en cualquier caso, el INE, por último, se encargará de aplicar los tratamientos finales, previamente comunicados. El escaso componente manual de estas últimas depuraciones facilitará mucho la consecución de la reducción de plazos que se persigue.

El objetivo principal es que el producto más importante y conocido de unos Censos, a saber, las cifras de población, vea la luz a mediados de 2002, y que el resto de productos de difusión censal vayan saliendo a lo largo de 2002 y 2003, de manera que, a partir de finales de ese año, una vez cumplido el programa básico de difusión censal, todos los esfuerzos puedan concentrarse ya en sacar el máximo provecho al enorme potencial de los datos censales.