

ASPECTOS RELEVANTES DE LA ORGANIZACION DE SERVICIOS DE PROCESAMIENTO ELECTRONICO DE DATOS DE CENSOS Y ENCUESTAS EN AMERICA LATINA

*Abel Packer
(CELADE)*

RESUMEN

Los avances recientes en el procesamiento electrónico de datos han cambiado el panorama de la producción y análisis de estadísticas de censos, encuestas y registros, al que habrá que adaptarse para mantener y aun elevar el nivel de eficiencia alcanzado por las oficinas nacionales de estadística de la América Latina.

Para alcanzar ese objetivo parece imprescindible contar con personal muy calificado, lograr un acercamiento entre usuarios y productores de datos, incrementar el uso de paquetes o sistemas de programas, implementar oportunamente las nuevas metodologías, documentar apropiadamente datos y procesos y mantener estrecha cooperación con instituciones similares para el intercambio de experiencia.

<PROCESAMIENTO DE DATOS> <COMPUTADORA>
<DATOS ESTADISTICOS>

ELECTRONIC DATA PROCESSING SERVICES OF
POPULATION CENSUSES AND SURVEYS IN
LATIN AMERICA

SUMMARY

Recent progress made in the field of electronic data processing has brought about significant changes in the production and analysis of census statistics, surveys and registries. National statistical offices in Latin America will have to assimilate these changes in order to maintain and even increase their level of efficiency.

Several requirements appear to be indispensable to attain this objective, among which mention should be made of highly qualified personnel, a better understanding between users and data producers, increased use of programme packages and systems, timely implementation of new methodologies, appropriate data and process documentation, and a close cooperation with similar institutions for purposes of experience exchange.

<DATA PROCESSING> <COMPUTER> <STATISTICAL DATA>

INTRODUCCION

Los avances observados en el campo del procesamiento electrónico de datos durante los últimos años han introducido una nueva dimensión en la producción y el análisis de estadísticas originadas en censos, encuestas y registros. El uso creciente de diferentes fuentes de datos y el desarrollo de técnicas y métodos de análisis aplicables a los estudios de población y a las estadísticas sociales en general, contribuyeron al surgimiento de un campo bien definido dentro de la computación electrónica: el procesamiento electrónico de datos estadísticos.

En estas notas se pretende destacar y comentar algunos de los aspectos más relevantes que conforman las condiciones básicas para la adecuada explotación de los computadores en el procesamiento de datos estadísticos de censos y encuestas. Las ideas que se expresan son producto, fundamentalmente, de la experiencia del CELADE en el procesamiento de datos estadísticos, particularmente en los trabajos de asesoría a las oficinas nacionales de estadísticas. Otras han sido extraídas de la bibliografía que se cita al final de este artículo.

Desde el punto de vista de los interesados en el procesamiento electrónico de datos estadísticos —en especial las oficinas nacionales de estadísticas y censos— el desafío que se presenta en el futuro inmediato (censos de 1980) es la puesta en marcha de sistemas que hagan del computador una herramienta realmente eficaz para la producción y análisis de datos estadísticos, en vez de constituir una fuente de problemas, atrasos, pérdidas económicas y deterioro de la información. No existen para ello soluciones mágicas, pero la toma de conciencia del problema, como asimismo un análisis exhaustivo de la situación actual y de la experiencia pasada, en conjunto con una clara definición de objetivos, representan un buen inicio.

Al identificar el procesamiento de datos estadísticos como un campo bien definido dentro de la computación, nos basamos en el hecho de que, descontadas las particularidades de cada fuente de datos, los procesos se componen siempre de las mismas etapas: *a)* el ingreso de los datos a algún medio de almacenamiento accesible al computador, *b)* la estructuración de las unidades de estudio, *c)* la limpieza de los datos y finalmente, *d)* la etapa repetitiva o interactiva de producción de las estadísticas. Sin embargo, es muy común que en las instituciones productoras de estadísticas se plantee el uso del computador en algunas etapas como un problema nuevo, en el sentido de iniciar el trabajo de análisis y programación como si fuera una tarea con algoritmos y rutinas nuevas,

sin acumulación de experiencia propia o adquisición de tecnología y metodologías ajenas.

Por otro lado, es frecuente responsabilizar a los analistas y programadores por los problemas que se presentan en el procesamiento de datos, sin tomar en cuenta que éstos se deben más bien a una concepción inadecuada de la organización de estos servicios. La necesidad de usar computadores por parte de las oficinas nacionales de estadística resulta obvia, dada la imposibilidad de atender, sin ellos, la creciente demanda de estadísticas. Sin embargo, no se puede pensar que el simple acceso a un computador o el cambio a uno más poderoso eliminará los problemas inherentes a la producción y al análisis de datos estadísticos. Por el contrario, el uso de tecnologías avanzadas para solucionar estos problemas requiere que la organización del servicio de procesamiento de datos concuerde con los problemas a resolver.

Esta organización debería basarse en los siguientes aspectos:

- a) necesidad de contar con personal mejor capacitado en el uso de las amplias posibilidades del procesamiento de datos;
- b) acercamiento de los usuarios al procesamiento de datos, a fin de establecer la comunicación necesaria;
- c) uso de programas generalizados (paquetes o sistemas);
- d) implementación oportuna de las nuevas metodologías y sistemas;
- e) documentación de los datos y de los procesos;
- f) necesidad de establecer cooperación con otros centros similares para el aprovechamiento e intercambio de experiencias.

A continuación se analiza, de modo general, cada uno de estos aspectos. Su implementación en la organización de los servicios de procesamiento de datos estadísticos variará en tiempo y profundidad, de acuerdo a cada situación particular.

ASPECTOS RELEVANTES DE LA ORGANIZACION Y FUNCIONAMIENTO DE UN SERVICIO DE PROCESAMIENTO DE DATOS ESTADISTICOS

Los aspectos que aquí se analizan son imprescindibles tanto para el uso eficiente del computador como para la organización de un servicio

dinámico de procesamiento de datos, y por ello, sería conveniente que las instituciones pertinentes adoptaran las medidas que aseguren su implementación.

Además de los recursos de “hardware” y “software” que se pueden adquirir, la explotación del computador requiere una adecuada relación entre tres funciones distintas y casi siempre complementarias entre sí: *captación de los datos*, su *procesamiento* y el *análisis*. La captación de los datos y el análisis tienen contenido “sustantivo” y determinan, en última instancia, la entrada y la salida del computador. Al personal de computación le cabe la tarea de implementar los programas y procedimientos destinados a resolver los requerimientos de la parte sustantiva.

a) *Personal calificado*

No existen dudas en cuanto a la necesidad imperiosa de contar con personal calificado en las funciones de captación, procesamiento y análisis de datos estadísticos. Sin embargo, la calificación del personal de computación trasciende el simple dominio de las rutinas de cada función, pues incluye el conocimiento y la comprensión de los objetivos de cada una de las operaciones básicas que se aplican a los datos, desde la captación hasta el análisis. Este aspecto constituye el punto de partida para el establecimiento de un sistema eficiente en el uso del computador, ya que refleja en su funcionamiento la interrelación entre los encargados de cada una de las funciones.

En consecuencia, es indispensable que se asigne el máximo posible de recursos para la capacitación del personal de cada función, dirigiéndola siempre de modo que los nuevos conocimientos se reflejen en el desarrollo de todo el sistema.

La capacitación continua del personal debe concebirse más allá de la educación formal teórica, que desde luego es necesaria, pero insuficiente para obtener un cuadro de personal calificado. Este sistema se desarrolla en un nivel complementario —el de la práctica—, de la experiencia en el tratamiento de los problemas, casi siempre impostergables, que surgen a diario. Por esta razón es muy importante conservar vigente la experiencia pasada, como también mantener una actitud receptiva a nuevos métodos y técnicas.

En lo que respecta al personal especializado en computación, existe el problema de la alta movilidad ocupacional originada en la competencia entre las empresas privadas y el servicio público, que casi siempre se

resuelve a favor de las primeras. Es necesario, por lo tanto, mantener un sistema de trabajo que no se vea afectado por la frecuente pérdida de personal calificado, de tal modo que cuando ello ocurra, no se paraliquen los proyectos ni disminuya notoriamente la calidad del trabajo. Como criterio general, las tareas del procesamiento de datos deben realizarse con un intercambio permanente de experiencias entre los analistas y programadores, no depositándose en un solo funcionario el conocimiento y el manejo de programas y sistemas. Como norma, podría establecerse que cada especialista que reciba un curso de capacitación fuera de la institución en la que trabaja, se comprometa, por lo menos, a permanecer en la misma durante el tiempo necesario para transmitir a los demás integrantes del equipo los nuevos conocimientos adquiridos.

b) *Comunicación entre el personal de los servicios de computación y los usuarios*

La falta de comunicación o la comunicación deficiente entre productores y usuarios de datos y del servicio de computación, han constituido siempre una causa importante de las deficiencias del procesamiento de datos estadísticos. Este proceso de comunicación debe considerarse en dos sentidos.

Uno se refiere a las operaciones en las que se relaciona personal de diferentes funciones, pero que pertenece a la misma oficina que produce las estadísticas. En este caso, además de contarse con las condiciones ya descritas en el punto anterior, debería existir, en el plano directivo, una administración que establezca una estrecha cooperación entre los grupos. De no existir esta administración se corre el riesgo de subestimar o sobrestimar el papel del personal de procesamiento de datos (o el papel del computador como herramienta) en el desarrollo de los proyectos. Por ejemplo, cuando se proyecta realizar un censo o una encuesta, la comunicación debería empezar en una etapa temprana, con la participación del personal de los servicios de computación en el diseño de la boleta y en el diseño del sistema de códigos que se usarán en la clasificación de las respuestas, y también en las etapas de control de calidad de datos, limpieza de los mismos, y en el diseño del plan de tabulaciones. En otras palabras, el servicio de computación debe integrarse a la programación del trabajo y a los mecanismos de decisión, lo que contribuiría a aliviar muchos de los problemas que surgen habitualmente durante la etapa de procesamiento de datos, y a economizar tiempo en las tareas de diseño de los programas y su codificación simultáneamente con las etapas previas y durante la realización del censo mismo.

El otro sentido de la comunicación se refiere a la que debería existir entre los productores de estadísticas a nivel gubernamental y los usuarios externos, constituidos principalmente por centros de investigación o personas interesadas en la información estadística.

La falta de diálogo puede conducir a la producción de datos que no llegan a utilizarse y a la omisión de otros que serían necesarios para determinados estudios, o a la presentación inadecuada de los resultados. Este problema ha sido presentado con frecuencia —y con preocupación— en reuniones de especialistas. Cabe anotar que en el pasado la consideración del problema se ha visto magnificada por las diferencias de formación de uno y otro grupo.

Partiendo del tipo de organización o la naturaleza de la institución, puede pensarse en dos tipos de soluciones. Cuando se trata de centros de investigación que son a la vez productores y usuarios directos de la información (generalmente encuestas) la tendencia actual es proporcionar a los usuarios un contacto directo con el computador a través del uso de paquetes de programas con lenguajes orientados. Ello tiende a eliminar dificultades, ya que es más fácil entregar a estadísticos, sociólogos, demógrafos, algunos elementos de computación, que pretender enseñar a los analistas y programadores los aspectos teóricos en que se basan las investigaciones requeridas por los usuarios.

Si se trata, en cambio, de una oficina nacional de estadística que realiza un censo o una encuesta a nivel nacional, los principales usuarios de los datos serán oficinas de salud, de educación, de planificación, sin perjuicio de que también haya usuarios de tipo individual. En este caso, fundamentalmente por razones administrativas o burocráticas, aún es difícil establecer mecanismos que permitan a los usuarios utilizar sistemas generalizados para obtener directamente del computador los datos que necesitan, además de aquéllos que habitualmente se divulgan en publicaciones oficiales.

Podría facilitarse la comunicación si se contara con algunos profesionales especializados en temas sustantivos, que tuvieran, a la vez, formación en procesamiento electrónico de datos. Estas personas servirían de vínculo entre los usuarios y la oficina de computación, actuando en las diferentes etapas del proceso de producción de la información estadística.

c) *Programas o Sistemas Generalizados*

El desarrollo de programas o sistemas generalizados para el procesamiento de datos estadísticos ha creado condiciones o posibilidades muy diferentes a las que prevalecieron hasta fines de la década del 60, cuando se solucionaba cada problema específico con la elaboración de un programa específico, válido por una sola vez. Tal práctica es, hoy en día, inaceptable como norma general, ya que presupone que la solución de problemas similares pasa por las etapas de análisis, codificación de un programa en algún lenguaje de alto nivel (COBOL, FORTRAN, etc.), o en lenguaje de máquina, pruebas del programa y finalmente su ejecución. Esta práctica demanda gastos y tiempo excesivos para producir los resultados. El problema se agrava si se considera la ausencia de personal calificado o la falta de comunicación entre el personal de servicio de computación y los usuarios.

Los paquetes o sistemas generalizados ofrecen un lenguaje orientado a la solución de un problema, de modo que la programación se limita a la especificación de los parámetros que definen el problema, haciendo innecesaria la programación de rutinas y algoritmos.

Actualmente existen gran cantidad de "paquetes" de programas o sistemas orientados al procesamiento electrónico de datos estadísticos, que permiten agilizar notablemente las tareas de programación, obteniéndose resultados basados en algoritmos ya probados. Las oficinas nacionales de estadística, al igual que los centros de investigación, deberían estar al tanto de los avances logrados en este campo, para obtener e instalar aquellos programas que se adapten a sus necesidades y sean compatibles con sus equipos.

Al adquirir un paquete de programas debe conocerse previamente sus posibilidades de uso y la facilidad o dificultad de transferencia de una instalación a otra, así como de la calidad de su documentación, expresada en manuales de uso, manuales de operación, guías para el programador y ejemplos de aplicación. Además debería contarse, en lo posible, con los programas-fuente para efectuar las modificaciones deseadas.

d) *Oportunidad de la implementación de los sistemas.*

La elaboración e implementación de un sistema o procedimiento relacionado con el procesamiento electrónico de datos estadísticos requiere, por lo general, más tiempo del que suele suponerse. Al estable-

cer plazos para estas tareas, los analistas y programadores muchas veces subestiman las dificultades que suelen presentarse durante las pruebas. Por ejemplo, en la realización de un censo o encuesta, se planifica una serie de fases o etapas de trabajo que deben desarrollarse armónica y sucesivamente, y el no cumplimiento del plazo establecido para una de ellas repercute negativamente en todas las que le suceden. El procesamiento de estos datos requiere gran cuidado y un trabajo prolijo, ya que se manipula una gran cantidad de datos, y debe evitarse la repetición de procesos con todos los datos. Es decir, antes de la ejecución final de un programa, éste debe estar completamente probado, con absoluta seguridad en los resultados que se van a producir. Una manera eficaz de probar los programas, en relación con las diferentes situaciones que pueden producirse, es procesar en primer lugar los datos de un censo experimental, o mejor aún, una muestra de adelanto de cifras que sea representativa de todo el país. Además de la ventaja que significa, para los análisis preliminares, el disponer a corto plazo de resultados a nivel nacional, la experiencia que se adquiere en este trabajo resulta siempre sumamente útil para el procesamiento de la información de todo el universo.

La situación deseable es tener todos los programas codificados y probados después de procesada la muestra de adelanto de cifras, de modo que, a medida que se vayan ingresando los datos por unidades geográficas, su procesamiento se limite a una simple —pero cuidadosa— tarea operativa.

En general, la oportunidad en la implementación de los recursos necesarios se verifica cuando hay conciencia de que el servicio de procesamiento de datos debe contar no sólo con las soluciones, sino también con experiencia en la solución de los problemas que se presentan en el desarrollo de un proyecto. Esto exige como condición una estrecha comunicación con los usuarios, con el objeto de poder definir con anticipación los recursos necesarios para atender los requerimientos futuros.

e) *Estándares y Documentación*

La necesidad de contar con estándares y documentación es reconocida por todos, aunque en la práctica no se le asigne la debida importancia. Dos hechos, el gran volumen de los archivos de datos que se manipulan en las oficinas de estadística y la exigencia real de volver continuamente a los datos, hacen de este aspecto uno de los más relevantes para el procesamiento de información estadística.

Deben existir estándares para la descripción de los requerimientos que el usuario haga en relación con los procesos de transformación de los datos, la especificación de los subconjuntos de los datos a que se aplican, la ponderación de estos mismos, la exclusión de categorías de las variables, el diseño de la salida, etc. Deben también existir estándares para la descripción de los archivos de datos, para los procesos, la programación y la operación.

La existencia de estándares facilita la elaboración de la documentación de los datos y de los procesos. La documentación debe elaborarse siempre de tal modo que una persona con conocimientos similares, que no haya participado en el proceso pueda entenderlo, corregir posibles errores y usar los datos sin que se presenten situaciones imprevistas.

En el caso de los censos se podrían destacar dos tipos de documentación: una, previa a la elaboración de los datos e independiente de esta etapa de trabajo, aunque fundamental para que ella se realice de manera satisfactoria; la otra, relativa a la etapa misma de procesamiento.

El primer tipo de documentación que se menciona corresponde a los manuales de recolección de datos, los códigos e instrucciones para su uso y, en general, todo el material que permita saber cómo se definió cada variable y cuáles fueron los criterios de clasificación usados en cada caso, cómo está estructurado el "registro", etc. Aun cuando estos documentos se producen, suele descuidarse la importancia que tienen para el procesamiento futuro de los datos y con frecuencia se introducen cambios, sin dejar constancia de ellos. Esto ocurre con mayor frecuencia en el caso de "preguntas abiertas" para las cuales se cambian o se amplían las clasificaciones que fueron establecidas *a priori*. El segundo tipo de documentación se refiere a la descripción escrita, con diagramas y tablas de decisión, de los procesos, programas y archivos de datos. Estos aspectos son de importancia crucial para el buen desempeño de un servicio de computación; en particular, la documentación interna de los programas y los informes de control de los procesos constituyen parte importante del trabajo de los programadores. A cada archivo de datos debe corresponder, obligatoriamente, además de la descripción técnica, el contenido, la fuente y los procesos que lo originaron.

La documentación en general del sistema de procesamiento de un censo debería estar enteramente preparada antes de iniciar la elaboración de los programas. Esto permitiría definir con exactitud los procedimientos iniciales e intermedios de modo que el archivo final contenga

todos los requisitos para su explotación. La documentación de los resultados de cada etapa se produce a medida que se desarrolla el trabajo, y los cambios que se produzcan en la documentación mientras se elaboran los programas deben quedar registrados, de tal modo que el procesamiento y la documentación, terminen simultáneamente. Si por alguna razón se retrasa la preparación de la documentación, ésta perderá su riqueza de detalles y, con toda seguridad, también parte de su indispensable exactitud.

f) *Cooperación horizontal*

Tal vez el punto de partida para entender y solucionar muchos de los problemas del procesamiento electrónico de datos en los países de la América Latina, sea analizar en qué medida ellos se deban a que toda la tecnología, tanto la referida a máquinas, (“hardware”) como a programas (“software”) es transferida a la región, en forma de productos acabados, desde los países desarrollados.

Dicha tecnología es originada y diseñada para su implementación en condiciones (recursos materiales y humanos) drásticamente diferentes de las existentes en los países en vías de desarrollo. Este hecho es aun más importante si se considera que esa tecnología ha sufrido, en el curso de apenas veinte años, un vertiginoso desarrollo, en el cual la América Latina no desempeña papel alguno. Los países de la región no han participado, ni en términos de definición ni de demanda, en el desarrollo de la tecnología. Es dable esperar entonces que al saltar etapas en el avance de esta tecnología, la adopción, adaptación e incorporación vengán siempre acompañadas de problemas.

Las soluciones a estos problemas trascienden los límites de acción de los organismos productores de estadísticas, para ubicarse en el plano de las políticas nacionales sobre informática. Sin embargo, estos organismos deben participar en las soluciones objetivas en relación con la producción de estadísticas a nivel nacional.

Puede pensarse, por ejemplo, en un esfuerzo común destinado a la creación de las condiciones materiales y de personal capacitado, al intercambio de experiencias y al desarrollo de proyectos cooperativos entre organismos nacionales de estadística, con el apoyo de organismos regionales e internacionales que sea necesario. Este intercambio puede establecerse de diferentes maneras: conferencias, visitas mutuas, adquisición y adaptación de “software” para resolver problemas comunes, intercambio de especialistas, etc.

Ayer y hoy, los organismos regionales de las Naciones Unidas han ocupado un papel destacado en la transferencia de tecnología para el uso de computadores en el procesamiento de datos estadísticos. Como consecuencia de la experiencia pasada, habrá que orientar ese papel en el futuro para generar e incentivar en los países la organización y el uso de servicios de procesamiento de datos estadísticos que tomen en cuenta los factores mencionados anteriormente.