

Distr.
RESTRINGIDA

LC/R. 940
6 de noviembre de 1990

ORIGINAL: ESPAÑOL

CEPAL

Comisión Económica para América Latina y el Caribe

X Conferencia Interamericana de Estadística

Aguascalientes, México, 13 al 16 de noviembre de 1990

DIFUSION DE LA INFORMACION ESTADISTICA */

*/ Este documento fue preparado por la División de Estadística y Proyecciones.
Este trabajo no ha sido sometido a revisión editorial.

90-11-1751

INDICE

	<u>Página</u>
Introducción.....	1
Los canales de distribución.....	2
Organización institucional de la distribución de la información estadística.....	5
Conclusiones y puntos para el debate.....	6

INTRODUCCION

El presente informe tiene como objetivo establecer un marco de referencia para la sesión que sobre el tema "DIFUSION DE LA INFORMACION ESTADISTICA" sostendrán los Directores de Estadística de las Américas durante la X Conferencia Interamericana de Estadística que se llevará a cabo en Aguascalientes, México, del 13 al 16 de noviembre de 1990.

Es oportuno recordar que el tema mencionado es una preocupación permanente de la Secretaría de la CEPAL, quien ha designado a la División de Estadística y Proyecciones para llevar a cabo las tareas propias de la Institución en este campo así como también para promover y coordinar las actividades regionales en este tema, especialmente a través de las reuniones de los Directores de Estadística de las Américas y de los seminarios técnicos regionales pertinentes a este tema que se llevan a cabo bajo distintos proyectos ejecutados por la CEPAL.

El enfoque sistemático de este tema se inició en la Reunión de Directores de Estadística de las Américas, que se llevó a cabo en Santiago de Chile del 23 al 25 de septiembre de 1987 (Documento de Referencia No. 2). En la Reunión de Directores de Estadística de las Américas realizada en Santiago de Chile del 26 al 29 de septiembre de 1989 se continuó enfocando el tema de acuerdo a la decisión tomada en la anterior Reunión de los Directores, en el sentido de dar un lugar permanente y relevante a este tema en las Agendas de sus reuniones (Documentos de Referencias Nos. 14 y 15).

Es conveniente hacer notar que los aspectos técnicos y los sistemas computacionales disponibles para llevar a cabo las tareas en este campo han sido enfocados específicamente en dos seminarios regionales, uno bajo el proyecto de "Apoyo a las actividades de los censos del 90" llevado a cabo en Santiago de Chile en septiembre de 1989 sobre "Acceso de los usuarios a la información censal" y el otro bajo el proyecto "Estadísticas para el desarrollo. Sistemas Estadísticos Computacionales Generalizados" realizado en Cuernavaca, México, a fines de noviembre de 1989 sobre "Bases de Datos y difusión computacional".

Este trabajo presenta una revisión de los documentos ya mencionados y considerando algunas de las recomendaciones de los seminarios indicados propone también un conjunto de puntos para ser discutidos en la sesión que los Directores de Estadística sostendrán sobre este tema.

En la elaboración de este documento se ha contado, al igual que en los anteriores, con la colaboración del experto Sr. Raimundo Beca, bajo las mismas condiciones en aquellos establecidas.

LOS CANALES DE DISTRIBUCION

Siguiendo la organización de la presentación adoptada en el documento original enfocaremos, en primer lugar, los medios tecnológicos disponibles para la difusión de la información estadística.

En el Documento de Referencia No. 15 ya mencionado en la introducción, presentado a la Reunión de Directores en septiembre de 1989, bajo el punto "PRINCIPALES EVOLUCIONES TECNOLOGICAS", se expuso rigurosa y exhaustivamente los recursos tecnológicos de reciente aparición, resaltándose que esa situación se mantiene hasta esta fecha sin grandes cambios. En esta ocasión dedicaremos nuestra atención a sólo dos importantes medios : Bancos de Datos y Sistemas de Información Geográfica (SIG).

En relación a los Bancos de Datos puede afirmarse categóricamente que a la luz de la experiencia internacional, especialmente en los países más desarrollados en el campo del procesamiento computacional de datos estadísticos, este medio se ha constituido en el más apto y eficiente para atender tanto los requerimientos de procesamiento propios de las Oficinas de Estadística como particularmente para el acceso de los usuarios no especializados en computación a datos estadísticos, lográndose de este modo satisfacer el objetivo más importante de la difusión como es el cubrir computacionalmente el segmento de requerimientos de los especialistas temáticos.

Debe notarse que los bancos de datos realmente útiles son aquellos que permiten su acceso en línea ("on-line"), amigable o fácilmente, ya sea en modo local o remoto, dando al usuario la posibilidad de seleccionar el subconjunto de datos que es de su interés, los cuales pueden ser bajados ("download") a su propio equipo o eventualmente procesados en el equipo donde reside la base de datos utilizando paquetes generalizados de amplia difusión en el mercado.

En relación al "hardware" se puede disponer actualmente no sólo de los tradicionales poderosos computadores conocidos como "mainframes" sino también de los equipos microcomputadores utilizados de manera independiente o conectado en red local ("LAN"). La relativa nueva línea de equipos que está tomando una gran importancia está definida por las arquitecturas conocidas como abiertas, no propietarias, bajo el sistema operativo UNIX.

Sin lugar a dudas, las redes locales o los equipos UNIX, con su cada vez mayor potencia y capacidad, su gran conectividad local o remota, la disponibilidad de sistemas estadísticos computacionales generalizados muy amigables, sumados a la clara tendencia de menores costos, están llamados a ser los recursos más idóneos para la difusión de datos estadísticos "on-line".

En relación al "software" puede asegurarse ya que los paquetes relacionales disponibles en el mercado comercial, con lenguajes SQL, no resuelven eficientemente los problemas inherentes al procesamiento de datos estadísticos en su generalidad.

Con el fin de analizar con objetiva rigurosidad los tipos de paquetes disponibles, clasificaremos los datos estadísticos en tres grandes categorías : (a) censos y encuestas (b) series de tiempo y cuadros estadísticos (c) estadísticas de registro.

La categoría de censos y encuestas, definitivamente, requiere para su eficiente procesamiento la organización de los datos bajo la estructura conocida como transpuesta. Esta estructura de datos ha sido ya tratada satisfactoriamente en los equipos "mainframes" con paquetes especializados desarrollados ad-hoc por Oficinas de Estadísticas de algunos pocos países. El ejemplo más conocido, RAPID de Statistics Canada. Enfrentados al nuevo mundo de los microcomputadores y de las arquitecturas abiertas ya se está trabajando en la misma línea de RAPID. Una vez más, Statistics Canada, ha tomado el liderazgo en este aspecto a través de su sistema TRAFIC que muy pronto estará operativo. En el mundo de los microcomputadores bajo DOS está actualmente disponible el conocido y difundido sistema REDATAM.

La categoría de series de tiempo y cuadros estadísticos exige un sistema que administre y procese objetos. Existe un gran desarrollo en este campo, inclusive en el mundo comercial, con los sistemas conocidos como "Object Oriented Data Bases". Sin embargo, en las Oficinas de Estadística de varios países, se tiene una importante experiencia y desarrollo en este campo, mucho antes que en el mundo comercial. Como ejemplo, mencionaremos el sistema AXIS de Statistics Sweden, TEMPUS de INE España y recientemente URUCIB de la Oficina de la Presidencia del Uruguay.

Finalmente podemos afirmar que la categoría de estadísticas de registro pueden ser satisfactoriamente administradas usando sistemas relacionales con lenguaje SQL. El tipo de procesos y consultas de esta categoría de datos tiene un muy buen tratamiento bajo estos sistemas. Actualmente, en el campo estadístico, la función más relevante de un sistema SQL es su interconexión eficiente y simple con paquetes de análisis y tabulación estadística generalizados. Puede asegurarse que la "bondad" de un sistema relacional SQL está casi totalmente determinado por esta función. El otro aspecto que está siendo más y más exigido es la transportabilidad de las bases de datos y las aplicaciones, entendiéndose la transportabilidad como la independencia de las aplicaciones del equipo y sistema operativo utilizado. El sistema ORACLE, en este sentido, está tomando preponderancia en el campo estadístico ya que es operativo bajo equipos mainframes, minis y micros con distintos sistemas operativos, facilitando a los

usuarios el desarrollo y explotación de sus sistemas bajo distintos ambientes. No podemos dejar de mencionar, en el mundo de microcomputadores, el lugar que tiene el sistema DBase y sus compiladores. En los mainframes, ADABAS Y DB-2 tienen una importante presencia en el desarrollo de aplicaciones en áreas estadísticas específicas.

En cuanto a los Sistemas de Información Geográfica(SIG) podemos decir que su utilización está siendo paulatina, lentamente incorporada a las tareas de difusión y análisis de datos estadísticos.

Los SIG's, recordemos, por medio de una breve y válida definición, son sistemas de información que residen en bases de datos computarizadas que contienen simultáneamente datos temáticos y datos georeferenciales o espaciales, relacionados entre sí.

Es conveniente notar que desde el punto de vista de la definición precedente, un SIG, no es otra cosa que una bases de datos y un sistema de manipulación de la información contenida en esa base, características propias de cualquier base de datos, sean esos datos de naturaleza estadística o de cualquier otro campo de información. Más aún, si miramos a un SIG como la complementación de una base datos tradicional y una base de datos espaciales, que deben estar conectados lógicamente entre sí, concluiremos naturalmente que nos bastará disponer de una base de datos relacional con SQL para resolver todos los aspectos de su administración y procesamiento.

Si bien, en una simplificación inicial, un SIG es visualizable como similar a cualquier otra base de datos, tiene dos características fundamentales que la diferencian de otras bases y que al mismo tiempo la hacen compleja en su contenido y en sus sistemas de procesamiento.

Una de esas características está dada por el tipo de datos de entrada, ya que no sólo deben ser capturados datos numéricos y alfanuméricos sino también datos espaciales o geográficos que deberán ser captados desde fuentes tan tradicionales como mapas, datos analógicos, o desde sistemas sofisticados como "scanner" multiespectral o satélites, que ya brindan información digital. La fuente más usual de datos espaciales son los mapas que demandan técnicas de digitalización, relativamente sofisticadas pero siempre trabajosas de aplicar.

La otra característica discriminante de los SIG es que el tipo de datos de salida, casi siempre, son presentados sobre un sistema espacial, una superficie, un mapa usualmente. Esta función obviamente demanda una tecnología de uso no frecuente a nivel de equipo y del sistema de explotación. Más aún, si el SIG tiene como subproducto un sistema cartográfico, las demandas en equipo y sofisticación de sistema crece explosivamente, pues como es de

común conocimiento la producción computarizada de mapas está entre las técnicas gráficas/digitales más avanzadas.

En el campo de la difusión de datos estadísticos, los SIG's actualmente disponibles en el mercado están ya cumpliendo una importante y valiosa tarea de presentación gráfica de los datos estadísticos y no menos relevantemente están simplificando y asistiendo visualmente a tareas de planificación y análisis de factibilidad en distintos campos. En el campo estadístico, como ejemplo, pueden mencionarse la producción de mapas temáticos, de población, de indicadores socioeconómicos, de producción, etc. Las aplicaciones no estadísticas más relevantes caen en el campo del manejo de los recursos naturales, del diseño de vías de comunicación, de planificación urbana, sistemas catastrales, uso de la tierra y otros similares.

Los SIG's actualmente disponibles, sin considerar costos de creación de las bases de datos temáticas y espaciales, demandan una inversión mínima, en equipos y sistemas, de alrededor de US\$ 10.000, monto no elevado, accesible para la mayoría de los países de la región. Sin embargo, un equipamiento completo y sistema ("software") sofisticado que incluya una gran gama de funciones implica una inversión que bordea los US\$ 40.000.

ORGANIZACION INSTITUCIONAL DE LA DISTRIBUCION DE LA INFORMACION ESTADISTICA

Siguiendo con los lineamientos del documento original, a continuación, brevemente, se expondrán algunos puntos sobre los aspectos institucionales para la difusión de datos estadísticos.

En relación a las instituciones nacionales, usualmente los Oficinas o Institutos Nacionales de Estadística, debe mencionarse que existe una clara tendencia regional de asignar a estas instituciones no sólo funciones de productoras y distribuidoras de estadística sino también hacerse cargo de las políticas y estrategias de informática nacional. Puede aseverarse que el reunir en una misma institución las tareas estadísticas y de informática son de gran conveniencia. Si a esas funciones, como ya sucede en varios países, se agregan las responsabilidades en el área de geografía se alcanza una satisfactoria integración orgánica de estas tres áreas estrechamente relacionadas entre sí. Una reorganización y redefinición de objetivos en esta dirección parece ser una recomendable línea de acción en el mediano plazo.

En la eventualidad que un país cuente con una organización integrada de estadística, informática y geografía, naturalmente se establecerá una unidad, seguramente de alto nivel en el

organigrama, que tomará a su cargo todos los aspectos de la difusión de la información estadística.

La experiencia regional, observada en la mayoría de los Oficinas de Estadística nacionales, indica que es necesaria la organización de una Unidad de Difusión de Datos Estadísticos (UDDE). Los resultados de las discusiones mantenidas sobre este aspecto orgánico de las Oficinas de Estadística llevan a concluir que esa UDDE debe depender sin intermediarios de la Dirección General de la Oficina, que debe ser dotado de personal altamente calificado y que debe disponer de un equipamiento computacional adecuado, así como de sistemas de procesamiento computacional avanzados para satisfacer los requerimientos de una estrategia multimedia de difusión.

En relación a las instituciones privadas de difusión estadística es oportuno notar que actualmente se ha establecido sólidamente la comercialización de acceso a datos estadísticos por medios de conexión remota. Varios de los servicios existentes brindan la posibilidad de realizar el eventual procesamiento de sus datos en el mismo equipo que reside la base de datos. El uso de equipos de microcomputación y su relativamente bajo costo de conexión a líneas públicas conmutadas o de redes de transmisión de paquetes "Packet Switching Data Network", conocidos por la sigla PSDN, han dado una gran dinámica a este mercado. Debe notarse que existe una gran especialización en los bancos de datos accesables, observándose corrientemente bases de datos económicos, demográficos, comerciales y de servicios.

CONCLUSIONES Y PUNTOS PARA EL DEBATE

Al igual que en el documento original, como parte final, se presenta a continuación un conjunto de conclusiones y recomendaciones que se estiman relevantes para un debate sobre este tema.

Una vez más es oportuno resaltar que la Secretaría de la CEPAL seguirá insistiendo, a nivel regional, sobre la conveniencia y oportunidad de ir estableciendo lineamientos matrices para las políticas, estrategias y procedimientos técnicos para una real difusión de datos estadísticos valiosos disponibles en las Oficinas de Estadística nacionales.

De los trabajos presentados en anteriores Reuniones de los Directores, de los resultados de los debates sostenidos en esas Reuniones y de las conclusiones y recomendaciones surgidas en los seminarios técnicos regionales, pueden exponerse los siguientes importantes lineamientos matrices :

Es cada vez más evidente cuán imprescindible resulta vincular orgánicamente los procedimientos de producción, distribución y utilización de datos estadísticos. En otros términos, ninguna de las tres fases mencionadas pueden cumplir satisfactoriamente sus fines de manera aislada.

La gran variedad de soportes físicos para la difusión estadística actualmente disponibles conduce naturalmente a recomendar la adopción de una estrategia multimedia de difusión. Es obvio que una estrategia de esta naturaleza permite distribuir la información en el medio más adecuado para cada tipo de usuarios y de aplicaciones, con las naturales ventajas en cada caso pero como contrapartida con la correspondiente demanda de tecnología más elaborada por parte de las Oficinas de Estadística.

La estrategia multimedia arriba mencionada no invalida el acerto de que el medio más recomendable y óptimo para la difusión es la adopción de una estrategia de bancos de datos "on-line". Una vez más, debemos resaltar, que los bancos de datos, de acceso eficiente y amigable, permiten generar información sobre cualquier soporte de un modo muy simple y al mismo tiempo hacen posible la descentralización geográfica y la interacción con otros bancos de datos conformando una óptima red de información.

Es fundamental notar que tanto la estrategia multimedia como la de bancos de datos exige una normativa de amplia aceptación para fines de conectividad, identificación e intercambio de información. En esta misma perspectiva el uso de sistemas generalizados y ampliamente aceptados juegan un papel muy relevante para el establecimiento de una red de información accesible y procesable por todos los usuarios.

Es muy importante resaltar que la experiencia internacional enseña que ningún limeamiento matriz es viable sin una real reorganización y redefinición de objetivos institucionales. En este sentido, es prioritaria la organización de una unidad orgánica exclusivamente dedicada a la difusión de datos, que acerque a los usuarios los datos disponibles y simplifique las técnicas y sistemas para su procesamiento directo por los expertos temáticos. Por otro lado, para mantener una necesaria evolución en los recursos utilizados en esta tarea deberá favorecerse la convergencia con otras tecnologías como las telecomunicaciones y medios audiovisuales, estrechamente ligadas a las técnicas computacionales, evitando que estos esfuerzos y desarrollos sean derivados a otras áreas o unidades orgánicas diferentes de la propuesta.

En relación a los SIG's es oportuno, considerando los censos de población y vivienda del 90, evaluar su incorporación a los

sistemas de difusión y análisis de las Oficinas de Estadística de la región. Los esfuerzos e inversión que esta decisión implica estarán largamente compensados con los logros alcanzables, en especial si esta tecnología se incorpora conjuntamente con un sistema de cartografía digital. En el último seminario regional el CELADE informó que la versión REDATAM+ tendrá interfases con algunos de los más difundidos SIG's existentes en el mercado.

Finalmente, por expreso pedido de los participantes de los tres últimos seminarios regionales relacionados a temas de procesamiento de datos estadísticos, se informa a los Directores de Estadística de las Américas de la recomendación tomada por unanimidad por los expertos nacionales sobre la urgencia y conveniencia de establecer un sistema de correo electrónico entre las Oficinas de Estadística de los países de la región con el fin de lograr un ágil y oportuno medio de soporte técnico mutuo. La CEPAL podrá incorporarse a un sistema de este tipo sin dilación ni dificultades técnicas.