

Información de Base

Coordinador: Luis Carreño Piera

1. Los bancos de datos urbanos

Antonio Flos Bassols

Introducción

La Planificación urbana requiere tener acceso permanente a grandes cantidades de información estadística referida al territorio sobre el que incide. La falta de información de base ha constituido siempre una de las principales dificultades con que han tropezado los planes de urbanismo.

Tradicionalmente la información se ha obtenido, ya sea aprovechando encuestas realizadas por otros organismos, ya sea a través de los datos de censos almacenados por la Administración. De un modo u otro, el hecho es que el urbanista se encuentra frente a una información estadística deficitaria y difícil de utilizar. Las causas son múltiples: definiciones incompatibles de una misma variable en las distintas encuestas, niveles de agregación excesivamente altos, tratamiento insuficiente o mal definido de la dimensión espacial, falta de actualización, elevados costes, nivel muy bajo de fiabilidad (o nivel de fiabilidad desconocido), información almacenada en forma prácticamente imposible de manejar, etc.

La introducción de los ordenadores puede haber facilitado en parte algunos tratamientos, pero no ha impedido la mayoría de estos inconvenientes. Por otra parte, la mecanización de determinadas fuentes de información representa unos costes prohibitivos, a no ser que dicha información sea posteriormente aprovechable para múltiples finalidades de distintos organismos o entidades.

Los bancos de datos urbanos repre-

sentan en este contexto un intento de superación de esta situación utilizando todas las posibilidades de la técnica informática actual. Sin embargo, como veremos a continuación, un banco de datos constituye algo más que la aplicación de los ordenadores a los tratamientos estadísticos. A grandes trazos, se trata de construir una organización destinada a la recogida y difusión de información estadística referente a un determinado territorio. La imagen de esta organización es más amplia que la de un Centro de Documentación que suele trabajar sobre información elaborada y no integrable. El banco de datos desciende al nivel de los datos individualizados almacenados en un sistema integrado de ficheros sobre el que es posible realizar todo tipo de tratamientos y ofrecer salidas de información muy diversas: impresión de tablas, visualizaciones sobre pantalla, cartografía, etc. Los usuarios del sistema podrán efectuar interrogaciones a distancia y trabajar, ya sea en tiempo real (pregunta-respuesta instantánea), ya sea en tiempo diferido para los procesos no instantáneos.

Es lógico que un planteamiento de este tipo resulte sumamente sugestivo para todos aquellos que diariamente se ven obligados a trabajar con importantes cantidades de información estadística, especialmente los urbanistas. Sin embargo, como veremos, un banco de datos urbanos no se limitará a tener como usuarios las oficinas de planificación urbana.

Nuestro enfoque del banco de datos es el de una unidad que trabaja con datos agregados y no agregados, y, por lo tanto, no constituye un sis-

tema alternativo al de documentación automática, sino que queda perfectamente vinculado a él.

El banco de datos como sistema de información

Un banco de datos es un tipo determinado de sistema de información. Llamamos sistema de información a un conjunto formado por una base organizada de datos y un procedimiento (sistema de proceso de datos) que permita almacenarlos, tratarlos y tener acceso a ellos. Un banco de datos urbanos será un tipo de sistema de información con una base de datos de carácter urbano o, si se quiere, con datos útiles para la planificación urbana.

La base de datos se organizará en ficheros conservados sobre soportes tratables mediante ordenador. En principio los datos almacenados serán individualizados, es decir, representarán atributos o propiedades de cada una de las entidades de la población. Sin embargo, en general coexistirán dichos datos con datos agregados, obtenidos mediante encuestas o mediante explotaciones de los datos individualizados. En un banco de datos urbanos, la base de datos tiene, entre otras, la particularidad de que los distintos ficheros se hallan integrados merced a un tratamiento especial de la localización geográfica que permite ligar todos los datos del sistema entre sí.

El sistema de proceso de datos se

halla constituido por el "hardware" (1) y el "software" (2). Aunque este aspecto será tratado más adelante, conviene citar aquí que el sistema de banco de datos se distinguirá por una elevada utilización de los llamados sistemas orientados al utilizador que permiten el acceso a la información mediante un lenguaje sumamente sencillo que no requiere conocimientos previos de programación. Ello permitirá que los usuarios efectúen determinadas preguntas, al sistema, desde un terminal a distancia, ya sea para recibir una respuesta inmediata, ya sea para ordenar un determinado proceso.

La formación de la base de datos

El primer problema a considerar es el de determinación de las fuentes que facilitarán la información necesaria para crear los ficheros. Para ello vemos en primer lugar las soluciones adoptadas en los Estados Unidos en distintas experiencias de bancos de datos.

Lo que podríamos llamar primera generación de bancos de datos en los Estados Unidos pretendía poner a disposición de un determinado número de clientes, públicos o privados, una serie de informaciones estadísticas referentes a una área determinada (un distrito, una ciudad, un condado, etcétera). La recogida de la información se efectuaba a través de importantes encuestas. El resultado fue totalmente negativo: la mayor parte de estos bancos de datos no han llegado a funcionar nunca. Ello se debe de forma prácticamente exclusiva, al enorme coste que suponía la realización de tales encuestas, lo cual no sólo dificultaba la recogida inicial, sino también la puesta al día de los datos, y hacía muy cara la petición de cualquier información almacenada. Los organismos que intentaron llevar a cabo realizaciones de este tipo (Estado de California, ciudades de Alexandría y Denver, Distrito de Colombia, Condado de Santa Clara), gastaron todo su

presupuesto en dichas encuestas y olvidaron, por así decirlo, los aspectos de gestión y tratamiento de la información y las relaciones con los usuarios del "banco".

Actualmente, y de forma muy incipiente, asistimos en los U.S.A. al nacimiento de una segunda generación de bancos de datos, los cuales siguen apuntando a los mismos objetivos que los anteriores, pero pretenden, en cambio, constituir la base de datos a partir de ficheros ya existentes en la Administración Pública o en empresas privadas suministradoras de gas, agua, electricidad, etc. Además no pretenden ya una simple acumulación de ficheros, sino la integración de todos ellos mediante una perfecta correspondencia entre las entidades observadas, los campos, los códigos, etc.

El hecho decisivo que hace viables estos bancos es el de aprovechamiento del subproducto de actividades administrativas. Cualquier experiencia que intente realizarse en este campo deberá utilizar dicho subproducto como la fuente más rica en información y menos cara.

Por su planteamiento, estos bancos de datos de segunda generación aparecen estrechamente ligados a la Administración. En primer lugar, la Administración es la principal fuente de alimentación de la base de datos, a través del subproducto informativo de los procesos administrativos. En segundo lugar, un sistema de banco de datos permitiría a la Administración una racionalización de su sistema informativo gracias a la integración de la información de que dispone y un rápido acceso a la misma, evitando así el grave y costoso problema de su dispersión. A título de anécdota citamos el Ayuntamiento de New-York, que tiene repartidos en sus distintos departamentos más de 40 ordenadores con sus ficheros, la mayoría de los cuales son incompatibles. Por otro lado, la justificación del coste que representa la creación y mantenimiento de un banco de datos requiere su utilización masiva. En ello la Administración puede jugar un decisivo papel.

No debemos olvidar tampoco el importante papel que jugarán las grandes compañías suministradoras de agua, gas, electricidad, teléfonos, etc. La información contenida en sus ficheros es sumamente interesante y es conveniente aprovecharla para la formación de la base de datos. Pero hay que tener en cuenta que dicha cesión tiene lógicamente como contrapartida —aparte de un compromiso de control de difusión— el tener acceso a los demás datos del sistema.

Estas observaciones llevan a considerar la necesidad de prever una estructura del "banco" suficientemente ágil para satisfacer las peculiares necesidades de información no sólo de las oficinas de urbanismo, sino también de la Administración en general, y de otros posibles utilizadores (grandes compañías, servicios de estudios, oficinas de estudios de mercado, etc.). Sólo previendo los distintos tipos de utilización, se puede hacer viable la constitución de un banco de datos.

En la mayor parte de los procesos administrativos, los problemas se hallan muy definidos y la labor de creación se dirige especialmente a la mejora de los procesos antes que a la solución de los nuevos problemas o a la satisfacción de nuevas necesidades. Por lo tanto, un banco de datos proyectado para la Administración debería dirigirse más que nada a ofrecer una respuesta inmediata a problemas conocidos y, por lo tanto, preprogramables. Por el contrario, la labor del planificador plantea continuamente nuevas necesidades y nuevos enfoques. Se hallará menos interesado en conocer datos sobre entidades concretas que en realizar complejos tratamientos estadísticos, tratamientos que irán cambiando según los resultados de experiencias anteriores. Del mismo modo, así como para los procesos administrativos, el banco producirá principalmente documentos o informes de formato *standard* o visualizaciones en una pantalla de rayos catódicos, para los urbanistas no existen formatos *standard* para las salidas del ordenador. Estas pueden necesitarse tanto en forma de tablas, de gráficos o de mapas, distintos para cada caso.

Organización de la base de datos

Los datos que se almacenan en un banco se refieren a entidades con una propiedad común a todas ellas; la posibilidad de ser localizadas geográficamente. El tratamiento adecuado de la dimensión espacial constituye la base del funcionamiento del banco. En efecto, la base de datos se compone de dos tipos de ficheros: los ficheros de datos propiamente dichos y los ficheros geográficos. Estos últimos no contienen información estadística *per se*. Únicamente contienen la localización geográfica de los datos de los demás ficheros. Esto permite efectuar cruces de ficheros. Por otra parte son absolutamente impres-

1 y 2. Los términos "hardware" y "software" son de difícil traducción. "Hardware" significa la parte "dura" del sistema, es decir el ordenador; el "software" o parte "blanda" del sistema es el conjunto de programas que el sistema tiene incorporados, ya sea para regir su funcionamiento general, ya sea para facilitar la programación de aplicaciones concretas. El "software" suele ser suministrado por las propias empresas constructoras aunque cada vez es mayor el número de empresas productoras únicamente de "software" destinado en general a completar el "software" básico del sistema con vistas a aplicaciones especiales.



cindibles para poder realizar salidas de ordenador de tipo cartográfico.

Esta organización geográfica no quedaría clara de no efectuar una distinción entre "continente" y "contenido". Se consideran "continente" aquellas unidades estadísticas —distrito, manzana, inmueble— que, por tener una localización geográfica fija constituyen un marco de referencia a otras unidades que llamaremos "contenido". Estas son aquellas unidades estadísticas —personas, establecimientos...— cuya localización geográfica se efectúa indirectamente a través de un continente al cual se hallan asignadas. Así, una persona podrá estar afectada a varios continentes: lugar de residencia, lugar de trabajo... De este modo, a través de la relación entre "contenido" y "continente", y entre "continente" y localización geográfica, tendremos todos los datos del sistema relacionados entre sí y será posible efectuar todo tipo de tratamientos.

En lo que se refiere al nivel de agrupación de los datos almacenados, es conveniente plantearse la cuestión de las relaciones entre los bancos de datos y los sistemas de documentación automática. Corrientemente suelen distinguirse en base a que el primero trabaja sobre datos individualizados, mientras que el segundo lo hace sobre datos agregados y elaborados. Sin embargo, la distinción no es tan clara, puesto que, como veremos, los bancos de datos almacenan también a menudo datos agregados. Actualmente, están en marcha en Europa varias experiencias de este tipo, que ligan estrechamente los dos sistemas dentro de una unidad común de información estadística. Dado el carácter muy especializado de esta discusión, en nuestra exposición hemos prescindido de ella y hemos partido de la base de un sistema de información que hemos llamado banco de datos, que contiene información estadística de ambos tipos.

En principio es conveniente dar prioridad a la información individualizada; a las ventajas que ofrece para tratamientos estadísticos une el interés que posee para determinados servicios de la Administración. Sin embargo, un organismo de información estadística poseyendo únicamente datos individualizados no es ni práctico, ni barato, ni posible en muchos casos, ni conveniente en muchos otros.

En primer lugar, la información individualizada existente hoy en día, en la Administración o en otras fuentes, es muy limitada. Por otra parte, su recogida implica siempre una encuesta exhaustiva, la cual tiene un coste

excesivamente elevado para que su peso recaiga sobre el banco de datos.

En segundo lugar, tanto su coste de almacenamiento como su coste de explotación, resultan muy elevados, ya que los ficheros serán voluminosos y los tratamientos largos.

En tercer lugar, es mucho más barata y, por lo tanto, factible, la actualización de los datos agregados, mediante una encuesta muestral.

Finalmente, determinados estudios podrían efectuarse utilizando información elaborada derivada de explotaciones anteriores efectuadas por el banco.

En general, las ventajas del almacenamiento de datos agregados se reducen a una disminución en los costes y, en muchos casos, a una mayor rapidez de acceso a la información. En cambio, los resultados obtenidos serán menos ricos desde un punto de vista estadístico y menos adecuados a las necesidades de cada caso concreto.

La variada procedencia de los datos a almacenar nos lleva a un nuevo problema de organización: la compatibilidad de la información dada por las distintas fuentes. En general, la incompatibilidad puede ser de tres tipos:

- De codificación
- conceptual
- geográfica.

El primer tipo es el de más fácil solución, por lo menos a un nivel teórico. Es el caso en que un mismo concepto se halle representado en dos ficheros distintos por dos códigos distintos. El problema puede solucionarse a través de la transformación de dichos ficheros mediante tablas de correspondencia entre ambos códigos o de cada uno de ellos con un tercero, que es el del banco de datos. Aun cuando el procedimiento parezca sencillo, implica unos costes que pueden ser elevados, puesto que, dado el volumen importante de los ficheros, será necesario, entre otras cosas, invertir un buen número de horas-máquina.

En cuanto a los tipos restantes de incompatibilidad, la solución escapa ya del terreno de la informática para convertirse en un problema que debe ser resuelto mediante una coordinación *a priori* de las investigaciones estadísticas que puedan interesar al banco. Cualquier tipo de remiendo posterior implicaría condicionantes demasiado importantes a la validez de la información.

Finalmente, nos plantearemos el problema de la actualización. Esta no será posible si antes de la constitución del banco no se ha previsto un plan de adquisición. Este plan requiere:

a) la realización de un inventario de toda la información que es posible obtener de las distintas fuentes de alimentación del sistema.

b) El establecimiento de acuerdos con dichas fuentes que aseguren la continuidad del suministro.

c) La organización del sistema de modo que acepte nueva información que llegue bajo formas variadas.

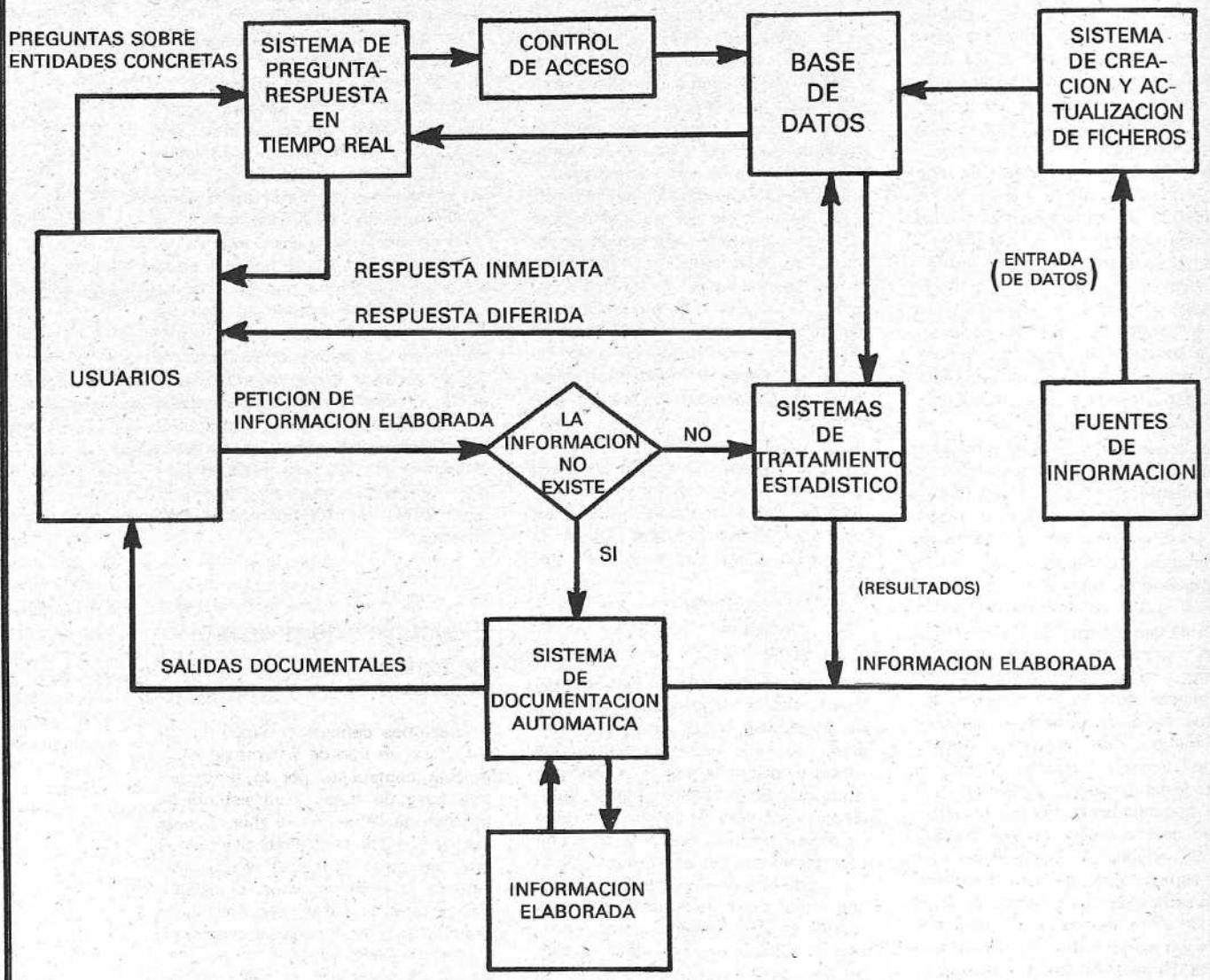
Antes de tratar del sistema de procesos de datos, mencionaremos dos cuestiones de funcionamiento general. Una es la necesidad de prever un mecanismo de "información sobre la información". Nos referimos a un sistema documental que ponga en conocimiento del usuario qué datos hay almacenados y qué tipo de información es posible obtener del sistema. Otra es la previsión de un mecanismo de control de acceso a los datos, es decir, se trata de que cada usuario pueda conocer únicamente los datos que le están permitidos. Este problema de control de acceso se da especialmente con los datos individualizados. Es posible crear un "software" que regule automáticamente esta cuestión.

El sistema de proceso de datos

Habíamos definido el banco de datos como un tipo de sistema de información, compuesto por lo tanto, por una base de datos y un sistema de proceso de datos. Hasta ahora hemos examinado los problemas planteados en torno a la formación y organización de la base de datos. El sistema del proceso de datos vendrá en parte definido por las especiales características de la base, es decir, de las "entradas" al sistema y, en parte, por las necesidades múltiples de los "usuarios".

Estas últimas pueden ir desde la simple consulta de una información individualizada a la resolución de complejos modelos de previsión. Sin embargo, aun con tan diversas necesidades, existen unos rasgos comunes a todos los usuarios en cuanto al modo de utilización. De todos, el más importante es la adopción de lenguajes orientados al usuario, que no requieren de éste ningún tipo de especialización en programación para obtener información del sistema. Estos lenguajes son posibles gracias a la aparición de unos *packages* "software" que llamaremos genéricamente sistemas de gestión de la base de datos (D.B.M.S.)

ESQUEMA DE FUNCIONAMIENTO DE UN SISTEMA DE INFORMACION ESTADISTICA



(3) que permiten el tratamiento de ficheros integrados (base de datos) con ficheros integrados (base de datos) con un grado de flexibilidad importante. Dichos sistemas han surgido de la necesidad de evitar la inversión en tiempo de programación y han sido lanzados al mercado por empresas especializadas.

Sistemas de este tipo son el "Mark IV" o el "Quick query" para interrogación y mantenimiento de ficheros, el I.D.S. de G.E. creado como extensión del lenguaje de programación, T.D.M.S. de S.D.C., orientado al acceso en

tiempo real, y G.I.S. de I.B.M., orientado al utilizador, constituido como extensión del sistema de operación para tratar una base de datos.

Los utilizadores de los bancos de datos pueden ser: órganos decisores de la Administración, servicios administrativos destinados a la tramitación más o menos rutinaria de expedientes, oficinas de planificación, compañías de estudios de mercado, servicios de estudios, grandes empresas, etc.

El sistema deberá prever todas las especiales utilizaciones que puedan requerirse. La simple consulta de información individualizada será propia

de aquellos departamentos de la Administración que lleven a cabo tramitaciones más o menos rutinarias de determinados expedientes y cuyas decisiones vengan dadas por el contenido del propio expediente o por las entidades relacionadas con él. Estos departamentos requerirán continuo y rápido acceso a los datos referentes a dichas entidades. Será para ello necesario instalar en estas oficinas un terminal -pantalla de rayos catódicos, impresora, etc- conectado al ordenador, a través del cual puedan realizar una serie de preguntas al banco de datos y obtener una respuesta rápida. Dado que estas preguntas responden

3. Data "Base Management Systems".