

LAS BASES DE DATOS. PANORAMA ACTUAL Y PERSPECTIVAS*

MANUELA VÁZQUEZ VALERO

Centro de Información y Documentación Científica (CINDOC). CSIC. Madrid

A partir de la segunda mitad del presente siglo, se ha producido un cambio sustancial en los medios de que dispone la humanidad para tener acceso al conocimiento publicado. Las revistas de resúmenes que, desde mediados del siglo XVIII se habían convertido en la herramienta básica para conocer los avances de la ciencia, habían ido diversificándose y especializándose debido al aumento espectacular de las publicaciones científicas, por lo que tanto su preparación como su consulta se habían convertido en un proceso lento y complicado.

La introducción de la informática en los procesos de composición de las revistas de resúmenes, a finales de los 60, dio lugar como subproducto a unas cintas magnéticas que almacenaban la información publicada en dichas revistas, si bien los registros sólo podían ser examinados de forma secuencial. El desarrollo de programas informáticos hizo posible, posteriormente, la búsqueda selectiva de información en dichas cintas magnéticas. Por otra parte el descenso del precio de los ordenadores, el descubrimiento de nuevos soportes de almacenamiento, y la evolución de las telecomunicaciones completaron los «ingredientes» que han dado lugar al desarrollo de las bases de datos, los medios actuales de acceso a la información.

Las bases de datos pueden ser definidas como colecciones completas de información relacionadas entre sí, que permiten la manipulación y la extracción de los datos que almacenan. Cuando se encuentran en soporte electrónico se consideran como servicios de información electrónica. Cuando se ofrecen en soportes unitarios, tales como diskettes, bandas magnéticas o discos compactos, se habla de ellas como productos.

Las bases de datos son sistemas de información constituidos por un conjunto determinado de datos y los dispositivos lógicos que facilitan su acceso y manipulación.

* Texto de la conferencia pronunciada en el Salón de Actos de la Biblioteca Nacional el 27 de octubre de 1993, organizada por ANABAD.

Desarrolladas bien entrada la segunda mitad del presente siglo (a principios de los años 70), su crecimiento ha sido muy rápido en todo el mundo y ha dado lugar al desarrollo de una industria específica.

1. LA CADENA DE SUMINISTRO

La industria de las bases de datos comprende una serie de sectores que corresponden a los eslabones existentes en la transferencia de la información (los datos) desde su producción hasta el usuario final. A diferencia de lo que ocurre con las publicaciones impresas, en las que entre el productor o autor y el usuario o lector sólo existen dos eslabones (el editor y el librero o vendedor), en el caso de las bases de datos hay bastantes más actores intermedios, aunque algunos de ellos pueden fusionarse o incluso no existir.

En el suministro de información a través de bases de datos pueden presentarse hasta siete eslabones:

- productor de información
- productor de la base de datos
- distribuidor
- operador de red
- «gateway» o pasarela
- contratista de información
- red de área local
- intermediario de información
- usuario de información

Algunos de ellos están relacionados con el contenido (productor de información), otros con la forma (productor de bases de datos y distribuidor) y otros con la comunicación (redes y gateways).

La situación se complica por el hecho de que algunos de estos papeles pueden fusionarse. Así, por ejemplo, el productor de información y el productor de bases de datos pueden ser la misma entidad; lo mismo puede ocurrir con el productor de bases de datos y el distribuidor; con la red y el gateway, etc. Es decir, pueden existir enlaces directos que circunvalan o fusionan los eslabones intermedios creando una compleja serie de relaciones entrelazadas.

La cadena de suministro es, por consiguiente, larga y compleja: larga porque tiene muchos enlaces y compleja porque se pueden combinar muchos de los papeles desempeñados por cada eslabón.

1.1 *Productores de información*

El *productor de información* es la entidad que realmente crea la información. Puede ser un editor que buscando una nueva ruta de marketing trata de vender su información electrónicamente. Puede tratarse de un departamento del gobierno, por ejemplo el Ministro de Comercio, que recopila estadísticas relacionadas con importaciones y exportaciones. Puede ser un hospital o un centro de investigación médica que registra información epidemiológica, etc.

Existen dos tipos de productores de información: editores que ya están publicando información impresa y organizaciones que producen la información como parte de su actividad o como subproducto de ella.

1.2 *Productores de bases de datos*

El *productor de bases de datos* es la entidad que transforma la información en base de datos. Puede ser un departamento de una empresa de publicaciones, una compañía de servicios informáticos o una firma de preparación de datos. Los productores de bases de datos pueden obtener la información con la que trabajan de los productores de información o bien pueden generarla ellos mismos. Por ejemplo, existen diversos productores de bases de datos (como la Institution of Electrical Engineers con su base de datos INSPEC) que examinan la literatura mundial en busca de artículos relevantes y almacenan sus resúmenes en bases de datos de referencia.

Los productores de bases se llaman a veces editores de bases de datos. Son editores en el sentido que ellos hacen públicas sus bases de datos. Algunos publican simultáneamente versiones impresas mientras que otros lo hacen sólo en forma electrónica. Los productores son responsables de la determinación del contenido de las bases de datos y de su producción.

Atendiendo a su procedencia institucional, los productores de bases de datos se clasifican en organismos oficiales, empresas o entidades privadas, instituciones sin ánimo de lucro y mixtos. Entre 1985 y 1992 han aumentado de forma continua los productores oficiales aunque su proporción relativa ha disminuido (Tabla I). Por otra parte, muchas de las bases de datos producidas por entidades privadas se basan en datos oficiales (por ejemplo, datos del censo de población). Es decir, algunos productores de bases de datos utilizan datos oficiales y después de estructurarlos de forma que aumente su valor informativo, los venden en forma de bases de datos comerciales.

TABLA I

BASES DE DATOS Y PROCEDENCIA DEL PRODUCTOR
(Porcentajes)

<i>Productor</i>	1985	1988	1989	1990	1991	1992
Organismos Públicos	21	20	17	17	16	15
Entidades Privadas	57	65	68	68	70	75
Inst. no lucrativas	11	13	12	12	12	9
Mixtas	11	2	3	3	2	1
<i>TOTAL</i>	100	100	100	100	100	100

Fuente: Gale Directory of Databases 1993.

La mayoría de los productores ofrecen sus bases de datos a organizaciones privadas que se encargan de su distribución. Hay productores de bases de datos que hacen también la distribución de las mismas.

1.3 *Distribuidores o hosts*

El distribuidor es una entidad que opera un gran sistema de ordenador central que puede almacenar una o varias bases de datos y las hace disponibles a través de las conexiones de telecomunicaciones a los otros eslabones situados por debajo en la cadena de suministro.

Los distribuidores son organizaciones que realizan una manipulación de valor añadido a las bases de datos y ofrecen servicios de acceso a las mismas, o distribuyen el producto en forma de CD-ROM.

El valor añadido que los distribuidores aportan a las bases de datos incluyen el software o lógico que facilita la búsqueda y dar servicio a los usuarios lo que incluye entrenamiento en el uso, servicios de difusión selectiva de la información y, a veces, suministro bajo demanda, de los documentos primarios que corresponden a los registros de las bases de datos bibliográficas.

Entre los más importantes se encuentran DIALOG, ORBIT, BRS, etc. que distribuyen sobre todo bases de datos textuales y en especial bibliográficas. En el caso de bases de datos de negocios, los distribuidores más importantes son el grupo WEFA, Reuters y DRI y entre los de bases de datos numéricas CAN/SND y STN International.

1.4 Operadores de red

La red es el lazo de comunicaciones entre el distribuidor y los eslabones situados más abajo en la cadena de suministro. Puede tratarse de la usual red telefónica conmutada (RTC) de la Compañía Telefónica o de una red de conmutación de paquetes. Puede ser también una red que incorpora otros servicios como el correo electrónico o el intercambio de datos electrónicos.

1.5 Pasarelas o gateways

El gateway es una entidad relacionada con una o más redes o distribuidores que provee un único punto de acceso a la información electrónica y a los servicios transaccionales.

Existen diversas definiciones del término «gateway». La más simple es que son puentes electrónicos mediante los cuales el usuario tiene acceso a diversos servicios de información estableciendo contacto solamente con uno de ellos.

Son, por consiguiente, dispositivos informáticos que pueden instalarse entre dos o más sistemas de acceso en línea de manera que se pueda ampliar el surtido de bases disponibles en uno de ellos, permitiendo la consulta de las bases de datos de los otros.

Hay varias clases de gateways. Algunas son también distribuidores de bases de datos en línea, o sea, tienen también bases de datos residentes en su propio ordenador. Otras, que no tienen bases de datos, pueden transferir las llamadas de los usuarios a los servicios en línea escogidos o pueden ofrecer un software que actúa como interface *frontal o amigable* que ayuda a seleccionar y usar adecuadamente las bases de datos.

Los sistemas de gateway pueden dividirse en:

Gateway simple en la que el usuario necesita un contrato y una contraseña para acceder al nuevo sistema y tiene que utilizar el lenguaje del sistema a que se conecte.

Gateway normal en la que el usuario no necesita ni contrato ni contraseña para el nuevo servicio en línea.

Gateway inteligente que dirige al usuario a diferentes servicios en línea sin necesidad de contratos ni contraseñas y permite usar un único lenguaje de búsqueda para todos los sistemas.

Los gateways entre distribuidores son algo más que enlaces de telecomunicaciones ya que implican convenios de cooperación comercial. Después de muchos años de competencia, los distribuidores han aceptado la cooperación, especialmente aquellos cuyas bases de datos se complementan.

1.6 *Contratistas de información*

Los contratistas de información son las entidades que se encargan de buscar información en nombre de los usuarios. Pueden buscarla accediendo en línea a las bases de datos o en bibliotecas; ya sean referencias o documentos originales, y pueden tener que integrar información procedente de distintas fuentes para satisfacer la demanda del usuario.

Los contratistas de información se encuentran en nichos de mercado específicos. Por ejemplo, el arrollador éxito de los servicios en línea de bases de datos de legislación en EE.UU. ha provocado el desarrollo de empresas especializadas en búsqueda de información en leyes, las cuales liberan al abogado en activo de perder tiempo en la terminal. El equivalente en Europa tardará en llegar, puesto que los servicios en línea de información legislativa están siendo aceptados más lentamente.

Igualmente, la creciente importancia de la búsqueda en línea de patentes y marcas ha supuesto que los agentes de la propiedad industrial hagan el papel de contratistas de información para su clientes más pequeños. Las grandes compañías químicas y farmacéuticas se pueden permitir tener sus propios especialistas (esto es, sus intermediarios de información), pero las pequeñas compañías utilizan los agentes de patentes y marcas como contratistas de información.

La tercera gran área para los contratistas de información son las noticias comerciales y financieras. Las nuevas compañías de especialistas en este campo han asumido el papel que antes tenían las agencias de recortes de prensa. Leen, buscan y reúnen noticias y las venden como servicio a sus clientes. Muchas de ellas, como por ejemplo Infomat en el Reino Unido, han llegado a convertirse en productores de bases de datos vendiendo el acceso a sus archivos a muchos usuarios.

1.7 *Redes de área local*

La red de área local es el dispositivo que provee las comunicaciones dentro de la organización. Usualmente proporciona conexiones con los sistemas internos, con los demás usuarios de la red (a través del correo electrónico) y con las fuentes de información externas.

Las redes internas han sido extensamente usadas en las empresas para el tratamiento de datos desde finales de los años 60. Casi siempre eran redes en estrella, lo que significa que tenían un ordenador central al que se conectaban todas las terminales u ordenadores subsidiarios. Si dos usuarios de la red querían comunicarse lo hacían a través del ordenador central.

Al ir entrando en escena cada vez más los mini-ordenadores y ordenadores personales, fue quedando claro que las redes tenían que proporcionar al-

guna forma de comunicación de punto a punto, sin tener que cargar la tarea de conmutación al ordenador central. Así empezaron a existir las redes de área local (LANs). A diferencia de las primeras redes en estrella las LANs tienen estructura de anillo o estructura en paralelo.

Tienen tres características destacables:

1. Al pertenecer a una única organización, su diseño y funciones pueden ser altamente especializados.
2. Cubren un área geográfica limitada, por tanto, su velocidad es muy alta en comparación con las redes públicas y con otras redes de larga distancia.
3. La tecnología de comunicaciones proporciona funciones tales como el correo electrónico, el acceso compartido a bases de datos internas y el acceso a servicios externos a través de un punto único de interface.

Esto último implica que de hecho una red de área local puede funcionar como gateway para sus usuarios, realizando las funciones de conexión telefónica, entrada de contraseñas y contabilización del uso de la información.

1.8 *Intermediarios de información*

Los intermediarios de información son profesionales que realizan búsquedas en línea para los usuarios finales. Resultan importantes para los distribuidores ya que son su principal mercado. La mayoría de los servicios en línea están diseñados para los intermediarios de información y son utilizados por éstos.

Muchos distribuidores han intentado prescindir de los intermediarios de información para poder vender directamente al usuario final, pero generalmente sus esfuerzos no han tenido éxito. Si los sistemas de búsqueda y recuperación se diseñan para el intermediario de información, serán demasiado complicados y difíciles para el usuario final. Si se simplifican para ayudar al usuario final, probablemente tendrán poca eficacia para los intermediarios de información, que hoy por hoy no pueden eliminarse.

Algunos distribuidores han intentado resolver este problema creando un sistema de acceso doble: un lenguaje completo de órdenes para usar por los intermediarios, y un sistema simplificado de menús dirigido a los usuarios finales. El problema de este sistema es que los intermediarios de información

pueden demostrar fácilmente que son capaces de hacer búsquedas más completas, baratas y mejores que los usuarios finales. El resultado es que los intermediarios de información siguen siendo los usuarios principales.

Así, la mayor parte de las estadísticas de uso de los servicios en línea, se basan en datos obtenidos de los intermediarios de información. Es evidente, sin embargo, que el mercado potencial de información está constituido por los usuarios finales.

1.9 *Usuarios finales*

Mientras que al hablar de libros y revistas hablamos de lectores, en el mundo de las bases de datos se habla de usuarios finales. Esta denominación se utiliza para diferenciarlos de los entes situados más arriba en la cadena de suministro (contratistas de información, intermediarios de información) que también son usuarios de información pero para suministrarla a otros.

Los usuarios finales constituyen para los servicios en línea un objetivo más difuso y difícil de identificar que el de los intermediarios de información. Los usuarios finales pueden apreciar el valor de la información para el desarrollo de su trabajo o el rendimiento de su empresa, lo que quiere decir que probablemente estarán dispuestos a pagar precios altos por una información que demuestre tener utilidad.

Por otra parte muchos se muestran reacios a aprender protocolos de comunicación, procedimientos de conexión con los distribuidores, lógica booleana y estructuras de las bases de datos, es decir, todo aquello que representa los conocimientos especiales de los intermediarios de información.

Si la tendencia de los servicios en línea es dirigirse al usuario final, tendrán que realizar fuertes inversiones para simplificar los procedimientos de acceso, ya que los usuarios finales no se sentirán atraídos por sistemas difíciles de usar.

2. TIPOLOGÍA DE LAS BASES DE DATOS

Existen unas siete mil bases de datos en el mundo a las que se puede acceder en línea. Es natural que en esta elevada cantidad de bases de datos, exista diversidad en cuanto a procedencia, tipo de datos que las constituyen, materias de que tratan, etc.

En consecuencia, hay muchas maneras de clasificar a las bases de datos.

2.1 *Por el carácter de la información*

Atendiendo al carácter de la información que suministran se habla de *Bases de datos fuente* y *Bases de datos de referencia*.

Las primeras proporcionan el dato original o texto completo de la fuente primaria, mientras que las segundas remiten a dichas fuentes primarias suministrando sólo una representación o resumen de ellas.

2.2 *Por la forma de presentación de los datos*

Otra posible clasificación es por la forma de presentación de los datos. Los datos pueden estar en forma de palabras, números, imágenes o sonidos. Atendiendo a esta presentación se habla de bases de datos textuales, numéricas, icónicas y de audio. La diferente presentación de los datos afecta también a la estructura de los ficheros y al software necesario para la búsqueda y recuperación, de forma que las técnicas de búsqueda también difieren considerablemente entre ellas.

Se han establecido varias subclases en las bases de datos textuales y numéricas. Las bases de datos textuales pueden dividirse a su vez en: bibliográficas, de patentes y marcas, directorios, diccionarios, de texto completo y otras. Las primeras bases de datos públicamente accesibles fueron las textuales bibliográficas.

Las bases de datos numéricas pueden ser subdivididas en: transaccionales, estadísticas, de series cronológicas, y de propiedades, como categorías principales.

Hasta mediados de la década de los 80 no existieron bases de datos icónicas y de sonido, de acceso público. Estas bases de datos han sido desarrolladas para microordenadores Apple Macintosh, IBM y NEXT. Con la popularización de los sistemas de hipertexto y de CD-ROM es muy probable que en los próximos años se desarrollen más bases de datos multimedia. Los CD-ROM multimedia combinan audio, vídeo y texto ASCII de manera que se puede recuperar simultáneamente imagen, sonido y texto e incluso permiten utilizar diferentes canales de audio para diferentes idiomas.

Todas las bases de datos se pueden clasificar de acuerdo con estas cuatro clases básicas y muchas pueden clasificarse de múltiples formas, es decir, textual-numéricas, textual-icónicas, audio-visuales, etc.

Sin embargo, hay dos tipos de bases de datos (las de servicios electrónicos y las de software) que se citan separadamente, pues aunque podrían encuadrarse en textuales y numéricas, dado que su empleo es diferente y tienen características especiales, es conveniente considerarlas como clases adicionales.

Los servicios de información electrónica tales como correo electrónico o conferencias electrónicas, contienen datos transitorios que no constituyen archivos. Sin embargo se encuadran con las bases de datos debido a que muchos distribuidores las incluyen entre sus productos.

Actualmente hay varios servicios comerciales que producen *bases de datos de software* accesibles en línea; así America Online suministra software Apple y Macintosh II; Byte Information Exchange, un servicio de conferencia electrónica, permite el acceso a software; Compu Help proporciona ayuda a usuarios ciegos e incluye entre sus servicios la teledescarga del software. Hay también productores que ofrecen bases de datos en CD-ROM que contienen software. Estas bases de datos contienen la documentación y codificación para los paquetes de software y son clasificadas en consecuencia como bases de datos de software.

TABLA II

CLASES DE BASES DE DATOS
(Cantidades y porcentajes)

Clases de BD	1985		1988		1989		1990		1991		1992	
	N.º	(%)										
Textuales	1926	(64)	3147	(69)	3409	(70)	4213	(72)	4661	(72)	6497	(71)
Numéricas	1084	(36)	1278	(28)	1250	(26)	1360	(23)	1422	(22)	1974	(21)
Icónicas (Vídeo)	—	—	16	(<1)	34	(<1)	98	(2)	151	(2)	358	(4)
De sonido (Audio)	—	—	1	(<1)	2	(<1)	16	(<1)	28	(<1)	109	(1)
Servicios electrónicos	—	—	101	(2)	136	(3)	178	(3)	179	(3)	193	(2)
De software	—	—	5	(<1)	10	(<1)	12	(<1)	57	(1)	91	(<1)
<i>TOTAL</i>	3010	(100)	4548	(100)	4741	(100)	5877	(100)	6498	(100)	9183	(100)

Fuente: Gale Directory of Databases 1993.

Considerando el total de bases de datos (Tabla II) se observa que las textuales son las más abundantes, seguidas de las numéricas. En parte esto es debido a que en las textuales se incluyen bases de datos de tipo directorio telefónico y de prensa así como las de datos de empresas.

Las bases de datos de imagen/vídeo se han multiplicado por 3,6 desde 1990 y por más de 10 desde 1989. El número de bases de datos audio también muestra un fuerte incremento pasando de 1 en 1989 a 109 en 1992.

En el subgrupo de bases de datos textuales (Tabla III) se observa que aunque el crecimiento de las bases de datos bibliográficas continúa moderadamente, su porcentaje con respecto al total va disminuyendo, debido princi-

palmente al incremento que han experimentado las bases de datos de texto completo y las de tipo directorio.

TABLA III

SUBCLASES DE BASES DE DATOS TEXTUALES

Subclase	1988		1989		1990		1991		1992	
	N. ^o	(%)								
Bibliográficas	1162	(37)	1223	(36)	1367	(32)	1425	(31)	1715	(26)
Patentes y Marcas	55	(2)	58	(2)	80	(2)	85	(2)	47	(<1)
De texto completo	1285	(41)	1412	(41)	1786	(42)	2040	(44)	3077	(47)
Directorios	613	(19)	707	(21)	952	(23)	1074	(23)	1611	(25)
Diccionarios	32	(1)	9	(<1)	23	(1)	32	(<1)	47	(<1)

Fuente: Gale Directory of Databases 1993.

2.3 Por procedencia geográfica

Las casi 7.000 bases de datos existentes en el mundo están producidas en 18 idiomas distintos y proceden de ocho grandes zonas geográficas: África, Asia y Oriente Medio, Australia, Oriente Lejano, Europa del Este, Europa Occidental, América del Norte y América del Sur (Tabla IV). En Asia y Oriente Medio se incluyen China continental y Formosa, India e Israel. En el Oriente Lejano se incluyen Hong-Kong, Filipinas, Japón, Singapur y Tailandia.

TABLA IV

BASES DE DATOS POR REGIONES GEOGRÁFICAS

Región	1991		1992	
	Número	(%)	Número	(%)
África	7	(<1)	7	(<1)
Asia	28	(<1)	25	(<1)
Australia	119	(2)	161	(2)
Oriente Próximo	155	(3)	171	(2)
Europa del Este	11	(<1)	12	(<1)
Europa Occidental	1473	(24)	1838	(26)
América del Norte	4424	(71)	4768	(68)
América del Sur	44	(<1)	16	(<1)
TOTAL	6261	(100)	6998	(100)

Fuente: Gale Directory of Databases 1993.

Las únicas regiones con más de mil bases de datos son América del Norte con 4.700 bases de datos y Europa Occidental con más de 1.800 bases de datos. A nivel de países, Estados Unidos es el país que ha producido más bases de datos; le siguen Inglaterra, Canadá, Francia, Alemania, Australia y Japón.

2.4 Por categorías temáticas

La mayor parte de los usuarios seleccionan las bases de datos en función de la materia de que tratan. Si por ejemplo desean información química, seleccionarán una base de datos de química, y será de importancia secundaria la forma o medio de la base de datos. El tipo de información que se desee puede dictar la necesidad de una base de datos de texto completo o una base de datos estadística.

Muchas bases de datos se pueden clasificar en dos o tres categorías temáticas. Cualquier categoría debe reflejar un porcentaje razonable del contenido de las bases de datos clasificadas en ella. En el caso de que el porcentaje sea muy pequeño, no se clasifica en dicha categoría.

El mayor directorio mundial de bases de datos existente (Gale Directory of Database) las clasifica en nueve grandes categorías temáticas, que son las recogidas en la Tabla V.

TABLA V

NÚMERO Y PORCENTAJE DE BASES DE DATOS POR MATERIA

<i>Materia</i>	1988	(%)	1989	(%)	1990	(%)	1991	(%)	1992	(%)
Biomedicina	433	(87)	586	(11)	651	(11)	690	(11)	728	(9)
Ciencias Exp. e Ing.	1184	(23)	996	(19)	1154	(19)	1210	(19)	1492	(19)
Ciencias Sociales	460	(9)	393	(8)	418	(7)	453	(7)	447	(6)
Derecho	441	(9)	447	(9)	531	(9)	574	(9)	885	(11)
Generales	310	(6)	327	(6)	416	(7)	450	(7)	700	(9)
Humanidades	84	(2)	184	(3)	216	(4)	248	(4)	314	(4)
Multidisciplinares	29	(<1)	335	(7)	368	(6)	366	(6)	296	(4)
Negocios	1815	(35)	1687	(33)	1956	(33)	2101	(33)	2624	(33)
Prensa	428	(8)	186	(4)	233	(4)	291	(4)	385	(5)
<i>TOTAL</i>	5175	(100)	5131	(100)	5943	(100)	6383	(100)	7841	(100)

Fuente: Gale Directory of Databases 1993.

Examinando esta tabla se observa que las bases de datos de negocios constituyen la primera categoría por su número, seguidas de las de ciencia y técnica y después por las de derecho (cuyo número se ha duplicado desde 1989). En cuarto lugar están las bases de datos biomédicas.

2.5 *Por la forma de acceso*

Otra manera de clasificar las bases de datos es en función de la forma de acceso o de soporte utilizado para su distribución. A pesar del incremento que ha experimentado el CD-ROM, todavía el 65% de las bases de datos producidas a nivel mundial son accesibles en línea, un 15% son distribuidas en CD-ROM, un 8% en diskette, 7% en cintas magnéticas y hay un 5% accesible en Batch (Tabla VI).

TABLA VI

MEDIOS DE ACCESO O DISTRIBUCIÓN DE LAS BASES DE DATOS

<i>Medio</i>	1989	1990	1991	1992
En línea	2711	3041	3200	4519
Batch	769	948	1014	320
CD-ROM	333	541	782	1088
Diskette	368	474	533	557
Cinta magnética	605	686	732	481
Portátiles	—	—	—	33
<i>TOTAL</i>	4786	5689	6261	6998

Fuente. Gale Directory of Databases 1993.

Algunas bases de datos son accesibles por más de una vía.

3. COSTES

Los costes cobrados a los usuarios por la utilización de las bases de datos en línea se pueden basar en dos elementos:

- Tiempo de conexión
- Registros recuperados

El tiempo de conexión era una parte muy importante de la tarifa aplicada en los primeros tiempos del acceso en línea. El ordenador era el elemen-

to más caro de la cadena, lo que significaba que era importante recuperar el coste de su uso. El coste de la información suministrada es el otro elemento de la tarifa.

Algunos distribuidores han establecido una cuota de suscripción o abono en sus tarifas, pero a nivel mundial esto es casi exclusivo de los servicios de información financiera de alto valor y de información de prensa.

En los últimos años se ha tendido a reducir la importancia del tiempo de conexión y a incrementar el precio de la información suministrada. Esta acción tendrá sin duda repercusión en el comportamiento de los usuarios finales. Si el tiempo de conexión encarece mucho el precio de una búsqueda, el usuario recurrirá a intermediarios de información cuya habilidad técnica garantiza un tiempo de conexión corto. Al reducirse, o incluso eliminarse, el costo del tiempo de conexión, el usuario prescindirá paulatinamente del intermediario, pues su falta de experiencia no dará lugar a una penalización excesiva del precio de búsqueda, y su conocimiento del tema le permitirá seleccionar eficazmente la información recuperada. Es posible que la competencia con los soportes en CD-ROM haya influido en esta tendencia.

4. EVOLUCIÓN MUNDIAL DE LA INDUSTRIA DE LAS BASES DE DATOS

La evolución mundial de la industria de las bases de datos puede expresarse en términos del número de registros producidos, número de búsquedas realizadas, cantidad de bases de datos existentes o del número de productores y distribuidores.

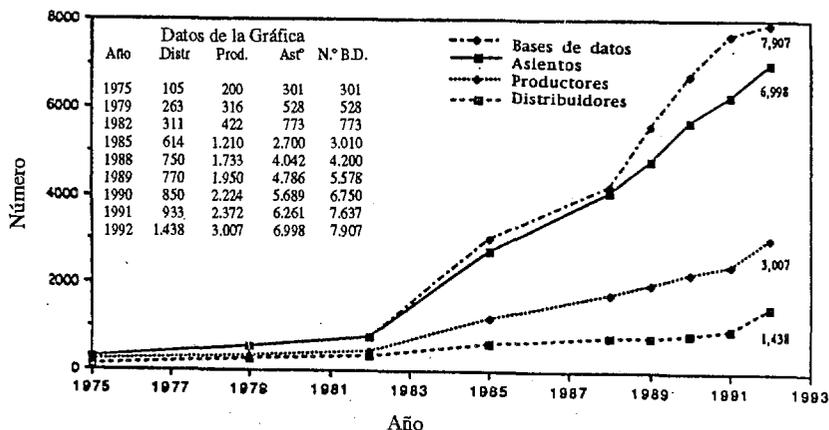
En el período comprendido entre 1975 y 1992 los registros incluidos en las bases de datos se han multiplicado por 87, pasando de 52 millones a 4.527 millones, mientras que el número de bases de datos se ha multiplicado por 26, pasando desde 301 a 7.907. Esto quiere decir que, además de aparecer nuevas bases de datos, las ya establecidas han incrementado su oferta de información.

El crecimiento del número de productores ha sido algo más lento, debido a que muchas de las nuevas bases de datos que entran en el mercado proceden de productores ya introducidos en él, habiendo pasado de 200 a 3.007 con un factor de crecimiento de 15 (fig. 1).

Algo semejante ocurre con los distribuidores, que han crecido muy lentamente debido a que cada uno ofrece numerosas bases de datos, algunos como DIALOG, varios centenares.

FIGURA 1

CRECIMIENTO DEL NÚMERO DE PRODUCTORES, DISTRIBUIDORES Y BASES DE DATOS



Fuente: Gale Directory of Databases 1993.

El crecimiento en el número de registros de las bases de datos no ha sido proporcional al crecimiento del número de bases de datos (fig. 2). Esto es, el tamaño medio de las bases de datos no ha permanecido constante. El crecimiento en el número de registros fue más bien lento hasta 1983, año en que se alcanzaron 310 millones de registros. De 1983 a 1984 creció más de tres veces (aproximadamente mil millones de registros); entre 1984 y 1987 esta cantidad se duplicó (más de dos mil millones de registros); y entre 1987 y 1992 se ha multiplicado por 2,19, alcanzando a finales de 1992 los 4.530 millones de registros.

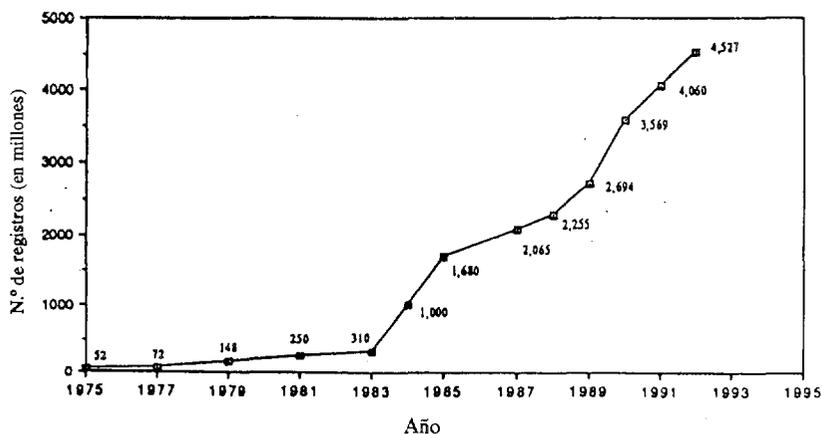
El tamaño medio de las bases de datos en 1975 era de 173.000 registros y pasó a ser de medio millón de registros en 1985. En 1992 el tamaño medio de las bases de datos era de 573.000 registros. Estos datos de tamaño-medio tienen un gran sesgo ya que hay 288 bases que tiene más de un millón de registros, y de ellas 10 con más de cien millones de registros.

El tamaño de cada registro también varía mucho, siendo el intervalo más frecuente entre 200 y 2.000 palabras. Los registros pueden ser citas, resúmenes, noticias, marcas, productos, datos estadísticos, imágenes, descripciones o listas de prácticamente cualquier cosa.

Probablemente el factor que más ha influido en el tamaño medio de las bases de datos ha sido el incremento de las bases de datos de negocios que comenzaron a mediados de la década de los 80. La introducción de grandes bases de datos de series cronológicas, de directorios telefónicos y de prensa a finales de los años 80, ha sido también responsable del incremento de dicho tamaño medio.

FIGURA 2

CRECIMIENTO DEL NÚMERO DE REGISTROS DE LAS
BASES DE DATOS



Fuente: Gale Directory of Databases 1993.

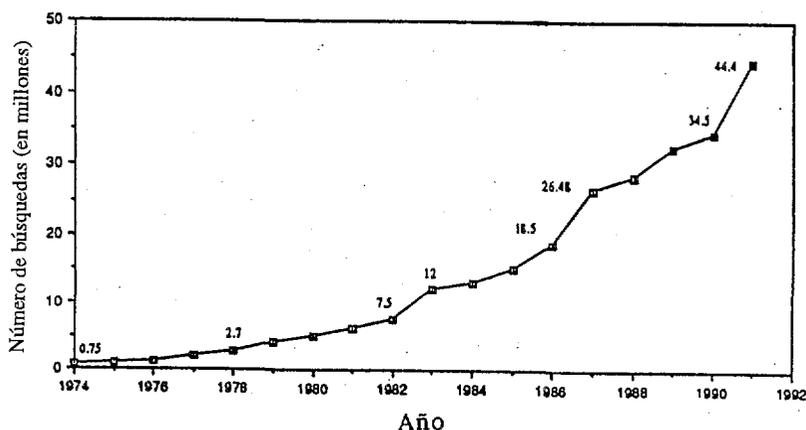
Entre las bases de datos más grandes se encuentran las bibliográficas seguidas por las de información sobre empresas, información sobre patentes, listines de teléfonos, datos químicos y demográficos, noticias de prensa, terminología y marcas comerciales.

Las estadísticas que muestran el incremento en el número de bases de datos, productores y distribuidores son indicativas del sólido establecimiento de la industria de bases de datos, pero el éxito real de la aceptación de los recursos legibles por computador y de la tecnología para acceder a ellos viene indicado por el crecimiento en el uso de bases de datos.

El uso puede medirse en términos del número de búsquedas o en términos de costes. Tales cifras son muy difíciles de conseguir para la industria de base de datos completa, y sólo existen aproximaciones. Algunos datos han sido producidos por empresas observadoras del mercado de información y por los principales distribuidores, pero estos muestran sólo una parte de dicho uso. Considerando únicamente Estados Unidos, se ha pasado de 750.000 búsquedas en 1974 a más de 44 millones de búsquedas en 1992, y esto teniendo en cuenta exclusivamente los datos suministrados por distribuidores (fig. 3). Si se sumaran a estos los relativos a transacciones bancarias directas, reservas de billetes de líneas aéreas y trenes, y sistemas de servicios a consumidores, el número evidentemente sería mucho mayor.

FIGURA 3

CRECIMIENTO EN EL NÚMERO DE BÚSQUEDAS EN ESTADOS UNIDOS
A TRAVÉS DE LOS PRINCIPALES DISTRIBUIDORES



Fuente: Gale Directory of Database, 1993.

Según las últimas estadísticas, que como hemos dicho son incompletas, la utilización de bases de datos en Estados Unidos es casi el triple de la de Europa y ésta a su vez más de tres veces la de Japón.

5. PANORAMA EUROPEO

Existe una marcada diferencia entre el patrón europeo de producción de bases de datos y el de Estados Unidos. En Europa, las bases de datos de referencia dominan la producción, mientras que en Estados Unidos lo hacen las bases de datos fuente. En conjunto, la Comunidad europea produce más bases de datos de referencia que los Estados Unidos y aproximadamente un tercio del número de bases de datos fuente. Esto demuestra la relativa inmadurez de la industria europea. A medida que en ésta se desarrollen las bases de datos de negocios, también se aumentará la producción de bases de datos fuente hasta llegar a una proporción de 3 a 1, similar a lo que ocurre en Estados Unidos.

Europa produce alrededor de un 12% más de bases de datos bibliográficas que Estados Unidos. Sin embargo estos producen un número mayor de registros bibliográficos, porque sus bases de datos bibliográficas son, por término medio, unas tres veces mayores que las europeas. En cifras absolutas los Estados Unidos han acumulado unos 200 millones de referencias bibliográficas mientras que la Comunidad Europea no llega a la mitad.

Por otra parte, hay que tener en cuenta que Europa es un conglomerado formado por más de 20 países en los que se hablan unos diez idiomas principales. Esta falta de homogeneidad unida a las relativamente caras telecomunicaciones, comporta que los servicios en línea en Europa tengan un coste superior al que tienen en Estados Unidos. Además, a esto hay que añadir que los distribuidores europeos están compitiendo con los norteamericanos que ya poseen el monopolio de suministro de bases de datos en sectores clave. De hecho muchas de las bases de datos europeas obtienen sus ingresos gracias a que son comercializadas por distribuidores de los Estados Unidos.

En comparación con los Estados Unidos, la industria europea de la información está en la infancia, aunque, como ya se ha indicado, la situación varía mucho de un país a otro (Tabla VII). El Reino Unido tiene la industria de información más fuerte de Europa. El hecho de tener el mismo idioma que Estados Unidos puede explicar en parte la causa de su desarrollo. Sin embargo, posiblemente sea el motivo principal el papel de Londres como uno de los más importantes centros financieros del mundo que ha motivado el establecimiento de diversos sistemas de información financiera (de los que Reuters es el más conocido) que suministran información a todo el mundo.

TABLA VII

PRODUCCIÓN DE BASES DE DATOS ONLINE EN LA CE
(En %)

Bélgica	2.3
Dinamarca	4.9
Francia	11.9
Alemania	19.7
Irlanda.....	0.2
Italia	11.5
Holanda	4.9
Portugal.....	0.3
España	7.6
Reino Unido	33.9
Comisión de la CE.....	2.6
<i>TOTAL</i>	100.0

Fuente: IMO Working Paper 90/5 (December 1990).

Entre las diez bases de datos más utilizadas en Europa, hay cinco producidas en Estados Unidos. Esto pone de manifiesto una elevada dependencia de los usuarios europeos de fuentes externas de información (Tabla VIII). En ciertos sectores, tales como en patentes y en medicina, esto es inevitable

ya que su esfera trasciende las fronteras nacionales. Sin embargo en el mundo de las bases de datos de negocios es importante que Europa sea autosuficiente porque esas bases de datos están estrechamente relacionadas con los intereses económicos de la Comunidad Europea y sus estados miembros.

TABLA VIII

BASES DE DATOS MÁS UTILIZADAS EN EUROPA

Chemical Abstracts	(EE.UU.)
Medline	(EE.UU.)
World Patent Index	(UK)
Inspec	(UK)
Compendex	(EE.UU.)
Biosis	(EE.UU.)
Pascal	(FRANCIA)
Predicasts	(EE.UU.)
Doma	(ALEMANIA)
Embase	(HOLANDA)

Fuente: BIS Mackintosh.

6. SITUACIÓN EN ESPAÑA

El panorama de la industria española de las bases de datos manifiesta un claro retraso frente a la media mundial. Este retraso se acusa en diferentes indicadores:

Así, en cuanto al origen institucional, la mayoría de las bases de datos (70%) son producidas por organismos públicos (como ocurría hace años en la industria norteamericana), cantidad bastante superior a la de la media europea (=50%), mientras que la empresa privada sólo produce alrededor del 15%. Otro 15% procede de instituciones no lucrativas.

Considerando el tipo de información, hay una clara preponderancia de las bases de datos de referencia (= 60%) sobre las bases de datos fuente. Entre las primeras hay una ligera mayoría de las bibliográficas frente a las de tipo directorio. En las bases de datos fuente las que predominan son las textuales y textual-numéricas.

Con relación al medio de acceso, casi un 80% son de acceso en línea; alrededor de un 7% se distribuye en forma de CD-ROM; 12% en diskettes y menos del 1% en forma de cinta magnética.

Una característica que pone en evidencia el claro retraso de la industria española de base de datos frente a la mundial es la falta de infraestructura de distribución comercial. España es el país europeo en el que la relación entre

distribuidores y productores es más alta. Es decir, muchos productores de bases de datos hacen el papel de distribuidores de las mismas.

En resumen, en cuanto a la producción de bases de datos el sector está dominado por los organismos públicos, con bases de datos pequeñas, con pocos usuarios, a las que se accede a través de distribuidores también pequeños y heterogéneos que comercializan inadecuadamente sus productos.

En términos de potencial en el mercado europeo será difícil que los servicios en línea españoles puedan competir en igualdad con los distribuidores de países con mayor tradición en estos servicios. Una posibilidad sería ofrecer datos con información sobre España que probablemente tendrían una buena acogida en el mercado europeo.

Por otra parte, España es el país mejor situado de Europa para proporcionar servicios en línea al gran mercado de América Central y Sudamérica y podría actuar como puerta de paso para otros servicios europeos que quieran llegar a ese mercado.

En cuanto al consumo de información, el incremento anual de servicios en línea extranjeros vendidos en España, es de alrededor del 12% (en comparación con el 5,7% mundial). Este es otro dato que manifiesta que en nuestro país hay varios años de retraso en cuanto a penetración de estos servicios.

7. MEDIOS ALTERNATIVOS Y COMPLEMENTARIOS

7.1 Videotex

Los servicios de videotex comenzaron a introducirse como un servicio experimental que permite a los usuarios comunicar de forma interactiva con bases de datos, mediante un conjunto de procedimientos normalizados, utilizando las redes de telecomunicaciones.

Creado a mediados de la década de los 70, el videotex es una extensión de los servicios informativos clásicos en tiempo compartido y está concebido para funcionar como un sistema de búsqueda de información autónoma o como una interface de empleo fácil para aplicaciones transaccionales y de bases de datos. Los usuarios pueden acceder a la información utilizando comandos que les son indicados y eligiendo opciones a través de un conjunto de menús estructurados, o por medio de otras técnicas sencillas que les permiten navegar en las bases de datos.

El videotex fue diseñado pensando sobre todo en el mercado del gran público, para simplificar el acceso de consumidores a los servicios de información y de transacción. En un segundo tiempo la experiencia adquirida en los servicios al gran público permitió elaborar servicios profesionales fáciles de utilizar.

El uso de videotex para acceder a grandes bases de datos presenta una serie de problemas:

1. Los terminales de videotex no son adecuados para la búsqueda intensiva en bases de datos: las pantallas contienen pocos caracteres y la definición es pobre.
2. La impresión de los resultados de las búsquedas es difícil.
3. El tiempo de respuesta es muy lento.
4. Es prácticamente imposible teledescargar y almacenar los datos para tratarlos posteriormente.
5. La estructura de las bases de datos de videotex se basa en páginas en lugar de en registros y campos, lo cual significa que impone al usuario un método de menús pesado, y la búsqueda por palabras es relativamente difícil.

Los servicios videotex más utilizados incluyen: directorios electrónicos de teléfonos, servicios bancarios, reservas de billetes, mensajería electrónica, telecompra y otros servicios transaccionales. Estos servicios pueden variar de un país a otro. A pesar de su oferta de servicios dedicados al gran público, con la excepción de Francia, en todos los países hay más usuarios profesionales que, lo que podríamos llamar, usuarios domésticos.

Los servicios videotex de los países de la CEE han creado gateways para unirse entre ellos permitiendo a sus usuarios acceder a videotex de otros países en condiciones razonablemente satisfactorias. No obstante, subsisten problemas relativos al idioma (cada país usa el propio), a la facturación y a las ayudas a los usuarios. Por otra parte, las terminales videotex no-ASCII son de una tecnología rudimentaria y no se adaptan a la distribución de la mayor parte de las bases de datos profesionales existentes, debido a los problemas antes mencionados.

Por otra parte, el progreso tecnológico en el campo del videotex permite ahora la transmisión de fotografías, aunque para ello los sistemas existentes tendrán que modificarse para hacer posible una mayor anchura de banda.

Según un estudio realizado por el Observatorio del Mercado de Información (OMI) basado en una encuesta entre usuarios, el 57% respondió que el videotex no es apropiado para la búsqueda de información profesional y sí lo es para información al gran público (92%).

En los Estados Unidos el videotex es considerado un sistema costoso y lento. Sin embargo su implantación en Europa sigue aumentando (constituye el 86% del mercado mundial) y se han establecido estrategias nacionales para favorecer su utilización primero en Francia y más recientemente en Italia.

También varios países fuera del marco de la OCDE ofrecen ya servicios videotex como Brasil, Malasia, Taiwan y África del Sur.

No obstante, el porvenir del videotex se presenta, cuando menos, dudoso y no parece aventurado predecir que jugará un papel relativamente pequeño como medio de acceso a bases de datos y como instrumento de desarrollo del mercado de bases de datos.

7.2 *Bases de datos en CD-ROM*

El CD-ROM es un método de almacenamiento que utiliza la tecnología láser tanto para grabar información en un disco como para leer dicha información.

El disco utilizado para los CD-ROM es el mismo que se utiliza en los discos compactos de alta fidelidad, es decir un disco de plástico de 12 cm. de diámetro. Ambos productos emplean las mismas tecnologías básicas de fabricación, de manera que el CD-ROM se aprovecha de las economías de escala que reducen su precio.

Como su nombre indica, el CD-ROM puede ser leído pero una vez fabricado ni se puede grabar sobre él, ni puede ser modificado.

La característica más destacada de los CD-ROM es su elevada capacidad de almacenamiento que permite almacenar en un disco del orden de 550 megabytes de datos, es decir, la información contenida en unas 250.000 páginas impresas o la imagen completa de más de 10.000 páginas. Debido a que tanto la grabación como la lectura se realizan mediante un rayo láser, no se produce contacto físico, ni en consecuencia desgaste, con lo que la pérdida o degradación de los datos es poco probable.

En un mismo disco se pueden almacenar imágenes y texto, bandas sonoras integradas con las imágenes o independientes de ellas y el software para su recuperación.

Desde su aparición en el mercado en 1985, el CD-ROM ha ido ampliando su implantación. Desde el principio se vio como un soporte ideal para el almacenamiento de obras voluminosas tales como enciclopedias, directorios, etc. Asimismo empezó pronto su introducción en el mercado de las bases de datos. Así el directorio de bases de datos Gale, correspondiente al año en curso, indica que 1.088 bases de datos fueron comercializadas en soporte CD-ROM en 1992.

Al empezar a ofertarse bases de datos en CD-ROM se pensó que afectarían negativamente a las consultas en línea. Sin embargo, aunque esto ha ocurrido en casos concretos, la industria de la información en su conjunto considera al CD-ROM como un medio complementario más que alternativo

a las bases de datos en línea. Debido a los específicos puntos fuertes y débiles de ambos sistemas, cada uno se considera adecuado para diferentes tipos de usuarios.

Por ejemplo, una importante limitación de las bases de datos en CD-ROM es que comercialmente no es rentable su actualización en intervalos menores de 1 mes: de hecho la mayoría ofrecen actualizaciones trimestrales. Por tanto la utilización de los CD-ROM será importante en sectores en los que la actualidad no sea esencial. La información obtenida en búsquedas en línea seguirá siendo imprescindible cuando se quiera obtener la más reciente información en una materia.

No hay cifras concretas sobre la reducción que las bases de datos en CD-ROM hayan podido producir en el uso de las bases de datos en línea. Hay datos que indican que en ambientes universitarios y también en algunas empresas se ha cortado drásticamente el uso de estas últimas.

Por otra parte se ha observado un flujo de usuarios del CD-ROM a las bases de datos en línea que aminora los efectos antes citados. El CD-ROM es un excelente medio de entrenamiento para las búsquedas en línea. Una vez se han familiarizado con ellas, muchos usuarios quieren completar su bibliografía con búsquedas en las mismas bases de datos en línea actualizadas o en otras sólo disponibles en línea.

Así hay datos que muestran que la rápida introducción del CD-ROM en las bibliotecas americanas ha dado lugar a que personas que nunca antes habían realizado búsquedas en línea se habitúen a su consulta y de forma particular se transforman en usuarios en línea.

La información financiera o bursátil cambia continuamente; por consiguiente, el amplio mercado constituido por bancos, analistas financieros, inversores, etc. seguirá utilizando las bases de datos en línea.

En general, en los centros en que se realizan muchas consultas a unas bases de datos muy concretas, sobre materias en la que *la actualidad del dato no sea imprescindible*, resultará económicamente ventajosa la utilización de bases de datos en CD-ROM. Por el contrario, cuando se hagan consultas esporádicas, o sobre materias muy diferentes, no estará justificada desde el punto de vista del coste la suscripción a bases de datos en soporte CD-ROM.

7.3 Soporte papel

Es conveniente destacar que las versiones impresas de las bases de datos continúan siendo una alternativa válida a las búsquedas en línea. Las grandes bases de datos bibliográficas por su gran volumen se consultan más fácilmente en línea o en CD-ROM, pero muchas búsquedas de información, por ejemplo sobre organismos o empresas, pueden ser más fáciles y baratas

con directorios impresos. Cuando existe barrera idiomática, este hecho aún es más evidente. Así, los directorios de empresas españolas accesibles en línea son relativamente poco consultados desde el extranjero, mientras que sus versiones impresas tienen un mercado exterior consolidado.

8. SUMINISTRO DE DOCUMENTOS PRIMARIOS

En los últimos años se está realizando un gran esfuerzo para proporcionar acceso a documentos completos por medios electrónicos. Así, por ejemplo, ADONIS suministra documentos completos de revistas biomédicas editadas por un consorcio de editores, los cuales producen un CD-ROM que contiene imágenes completas de los artículos de unas 500 revistas de ese sector. Los discos son adquiridos por las bibliotecas mediante una suscripción fija, pero después deben pagar al editor un *royalty* o canon por cada artículo reproducido a partir de dichos discos. ADONIS ha supuesto un éxito importante en cuanto a desarrollo técnico, pero su futuro comercial es todavía incierto. Hay varias razones para ello: en primer lugar la información es limitada, pensemos que se editan más de 10.000 revistas de interés biomédico en el mundo y de ellas unas 1.200 son consideradas fundamentales, frente a las 500 que tiene ADONIS. Por otra parte si consiguieran aumentarlas para que su cobertura se ampliara no podrían mantener su programa editor de suministrar un CD-ROM por semana, y por otra parte tendrían que ampliar aún más el número de editores socios del consorcio, aumentando los problemas que actualmente tienen.

Uno de los últimos productos que recientemente ha llegado al mercado español, desarrollado por la University of Microfilms y distribuido por la empresa LUSODOC, consiste en la combinación de bases de datos bibliográficas y bases de datos de imágenes, ambas en soporte CD-ROM.

La base de datos bibliográfica está constituida por la descripción bibliográfica, términos de indización y resumen de todos los artículos contenidos en una amplia serie de revistas de un tema o especialidad concreto.

La bases de datos de imágenes contiene la imagen de los textos completos de los artículos que integran la anterior base de datos.

Por consiguiente, este producto permite una búsqueda convencional por cualquier campo del registro, visualizar los resultados de la búsqueda y obtener el documento completo de las referencias obtenidas.

En el momento actual ofrece tres productos:

- Información financiera
- Publicaciones de IEEE/IEE
- Índice de Ciencias Sociales

El primero contiene la información de más de 800 revistas internacionales (de las que 300 son europeas) de economía, gestión y ciencias afines (a partir de 1989).

El segundo está formado por la totalidad de los trabajos publicados en las 80 revistas, 360 actas de congresos y alrededor de 500 normas editadas por ambos institutos de ingeniería.

Por último, el tercero es la versión a texto completo del Social Sciences Index, con información extraída de unas 300 revistas de ciencias sociales.

La suscripción a estos productos, cuya actualización es mensual, permite por tanto el acceso inmediato a un gran volumen de información a texto completo, fácil utilización del sistema y posibilidad de obtener copias ilimitadas.

Sus costes son bastante elevados pero también es cara la suscripción en papel de esos productos, y cuentan con el valor añadido de permitir búsquedas por cualquier campo o elemento que dicha información contenga, mediante un sistema de fácil manejo, al alcance de cualquier usuario poco experto; y la posibilidad de obtener cuantas reproducciones se desee, sin coste adicional.

9. CONSIDERACIONES FINALES

El mundo de las bases de datos es dinámico: cambia y crece cada año y no se ven síntomas de que el crecimiento se vaya a estancar. Los cambios en los 17 años en los que se dispone de estadísticas mundiales, son espectaculares. Los nuevos medios para la distribución y el acceso a las bases de datos han aumentado el potencial para atraer nuevos usuarios. En particular, el desarrollo y la distribución de bases de datos en CD-ROM ha incrementado grandemente el uso de las bases de datos en las universidades y centros de enseñanza superior, en los que el coste había sido con frecuencia una barrera para el acceso.

Desde el punto de vista de los usuarios, la irrupción de los gateways con su finalidad, aún no lograda totalmente, de conseguir un punto único de acceso a todas las bases de datos existentes, facilitará el acceso a profesionales diversos como científicos, abogados, etc., es decir, no solamente a los especialistas en información. El desarrollo de los sistemas de búsqueda llamados amigables y la voluntad de los productores de ofrecer productos orientados al usuario final, favorecerá este cambio.

La industria de bases de datos se encuentra en una situación interesante: por una parte tiene la experiencia de más de 25 años en la que basarse, pero por otra tiene ante sí unas inmensas perspectivas de crecimiento hacia las que mirar.

10. BIBLIOGRAFÍA COMPLEMENTARIA

1. COLLIER, Harry: *Strategies in the Electronic Information Industry. A guide for the 1990s*. Calne. Infonortics Ltd.
2. DURIEUX, B.: *On line Information in Europe*. EUSIDIC. 1992.
3. *Gale Directory of databases*. 1993.
4. *Gateways: the growing trend in database distribution*. Information Market Observatory report 89/6.
5. O'BRIEN, R. C.; CHANNING, M.: *The global structure of the electronic Information Services Industry*. Oxford Survey in Information Technology, 1986.
6. *Production and availability of online databases in 1989*. Information Market Observatory. Working paper 90/5.
7. RUIZ GONZÁLEZ, B.; GONZÁLEZ MOLINA, M. T.: *Catálogo de Servicios españoles de información electrónica ASCII*. Fundesco. 1991.
8. *Services d'information informatisés: Les aspects économiques et commerciaux du marché des bases de données*. OCDE. 1992.
9. WILLIAMS, Martha E.: *Highlights of the online database industry: Have ten years gone by?* Proc. of the 10th National Meeting. New York, May 9-11, 1989. Medford NJ 1989. Learned Information, pp 1-4.