

DATATUR. ALMACÉN DE DATOS PARA EL ANÁLISIS Y DIFUSIÓN DE LA INFORMACIÓN ESTADÍSTICA DEL TURISMO EN ESPAÑA

José Manuel Salinas González*

Resumen. Se describe en este artículo el almacén de datos turísticos creado en el Instituto de Estudios Turísticos (DATATUR). Se da una visión global del sistema tanto desde su perspectiva metodológica como técnica, haciendo hincapié sobre todo en su arquitectura y funcionamiento.

Abstract. In this article is described the Tourist Information Data Warehouse developed in the Instituto de Estudios Turísticos (DATATUR). It shows a global vision of the system, its methodological perspective as technology, emphasizing especially in its architecture and functionalities.

I. INTRODUCCIÓN

El Sistema de Información de Estadísticas Turísticas (DATATUR) es hoy una realidad en el Instituto de Estudios Turísticos (IET). Su desarrollo, se ha hecho a partir de las especificaciones realizadas por el propio Instituto sobre lo que debía ser un Gran Almacén de Datos Estadísticos sobre Turismo (1).

La construcción de DATATUR se ha realizado por medio de técnicas de Data Warehouse, por lo que es un sistema en constante cambio y por tanto nunca puede darse por cerrado su desarrollo, ya que debe evolucionar y mejorarse en el tiempo, para adaptarse a los cambios que la realidad turística requiere.

El objetivo inicial que se quería conseguir era proporcionar al Instituto de Estu-

dios Turísticos una plataforma desde la cual poder satisfacer las demandas externas e internas de información coyuntural turística, basada en un sistema que aunara la facilidad de uso por parte de los usuarios tanto como la simplicidad en su mantenimiento. El sistema actual cumple hoy en día con esas expectativas y ya es la herramienta de referencia obligada para proporcionar información externa a través de Internet, para CC. AA., Oficinas Españolas de Turismo en el extranjero (OET's) y público en general, e información interna para uso de los técnicos de negocio del IET.

El desarrollo del sistema constituye una experiencia pionera en el desarrollo de bancos de datos estadísticos con tecnología OLAP.

En los próximos apartados se explicará cuáles han sido los pasos en su desarrollo,

* Gerente de BG&S Online Consultores.
Especialista en el desarrollo de Data Warehouse con tecnología OLAP.

los problemas encontrados, la tecnología empleada y sus perspectivas de futuro.

Para poder comprender algunos aspectos que se tratarán en este artículo es necesario definir una serie de conceptos relacionados con la arquitectura o modelado de los datos.

1.1. Conceptos OLAP

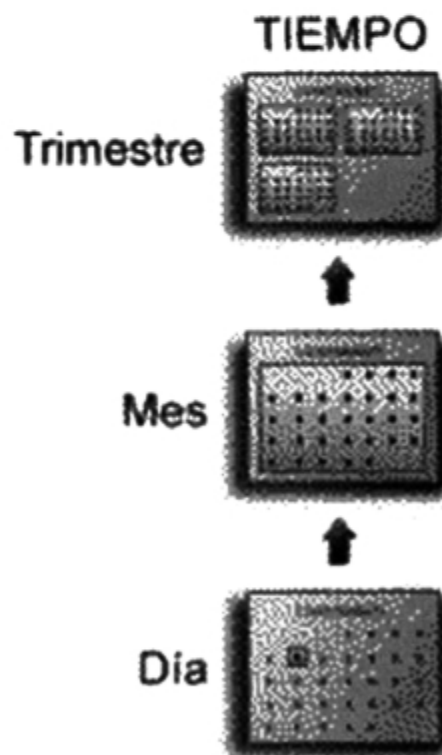
El modelado de datos tradicional y más ampliamente utilizado en las bases de datos relacionales, es el modelo de datos normalizado. Este modelo es ideal para el almacenamiento de datos transaccionales (operaciones realizadas, nombres de clientes, direcciones...), pero no está enfocado al negocio sino al proceso. Es, por tanto, un modelo de semántica débil desde el punto de vista de la decisión. El usuario tiene mucha autonomía, pero debe saber cómo consultar la información y además, este margen tan amplio de autonomía puede conllevar una pérdida de control, es decir, que cada usuario llegue a una cifra distinta según cómo extraiga la información, ya que las informaciones interesantes para el usuario no existen directamente.

Más enfocado a la decisión o al negocio es el modelado dimensional de los datos que deriva de los conceptos que han llevado a la emergencia de las bases de datos multidimensionales, llamadas bases OLAP.

El objetivo principal es el análisis de una serie de **indicadores**, así, para un comercial, los indicadores podrían ser el volumen de ventas o el porcentaje de consecución de objetivos, para un director el margen de

beneficio o el cash-flow y para un usuario de negocio del sector turístico el número de visitantes o el número de hoteles. Estos indicadores se analizarán a través de **dimensiones**. El tiempo es una de las dimensiones que encontraremos casi obligatoriamente en todos los modelos multidimensionales, permitirá seguir la evolución de los indicadores por años, temporadas, meses, semanas, etc. Otra dimensión o **punto de vista del negocio** por la que se podría analizar un indicador serían los productos, en el caso del comercial (ventas por producto) o el tipo de visitante (residente, turista, excursionista), en el caso de indicadores turísticos.

Figura 1
Ejemplo de jerarquía dimensión tiempo



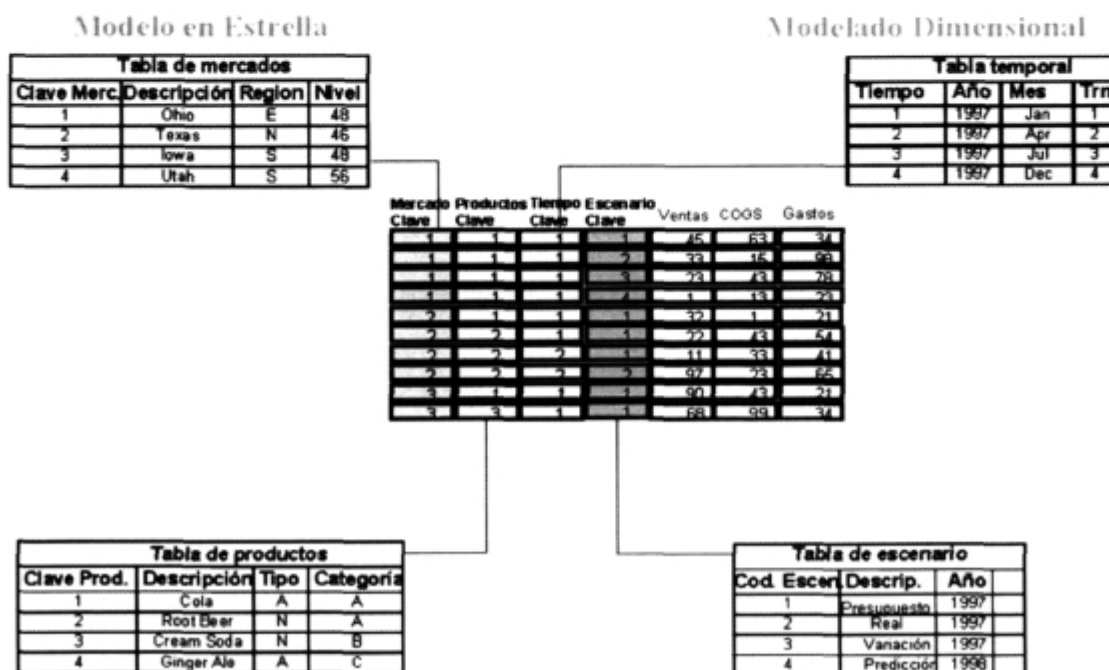
Los elementos de una dimensión se ordenan jerárquicamente, para luego poder «navegar» por ellos subiendo, «drill-up», o bajando, «drill-down», por la jerarquía. Por ejemplo, en una dimensión temporal, los días se unen en meses y los meses forman trimestres (véase figura 1).

Existen dos aproximaciones al modelado dimensional, que son el **modelo en estrella** y el **modelo en copo de nieve**. En ambos casos los indicadores se agrupan en una tabla central, llamada **tabla de hechos**. Una tabla de hechos agrupa todos los indicadores que comparten el mismo conjunto de dimensiones y que no pueden ser deducidos de otros indicadores.

En el centro de la estrella se coloca la tabla de hechos y a su alrededor las tablas de dimensiones. El identificador de la tabla de hechos es una clave múltiple compuesta por las claves de los elementos de cada una de las dimensiones de análisis (véase figura 2).

Un indicador está identificado, por tanto, en la tabla de hechos por todas las dimensiones que lo representan (para el caso de un viajero, el tipo de viajero, el tipo de alojamiento utilizado, el motivo de la visita, el país de residencia, etc.). La ventaja de este modelado es que la legibilidad para el usuario es muy alta, entiende perfectamente la finalidad del modelo y al estar orien-

Figura 2
Tabla de hechos



tado al tema o al negocio, el usuario sabe perfectamente cuáles son los indicadores.

El **modelado en copo de nieve** es una variación del anterior en el cual las tablas de dimensiones se subdividen en tantas tablas como niveles tenga la jerarquía de la dimensión, esto lleva a normalizar las tablas de dimensión y evitar las redundancias. En este modelado se ve más claro el concepto de jerarquía dentro de una dimensión.

Esta forma de trabajar tiene sus ventajas e inconvenientes en las que no entraremos por no ser la finalidad de este artículo, pero sí es interesante reseñar que no todo son ventajas, ya que, por ejemplo, este diseño está muy vinculado a la etapa de análisis, todos los indicadores que no se definan no pueden ser nunca consultados por los usuarios y se debe hacer un nuevo modelo o adaptar el existente para dar cabida a nuevos indicadores.

Una de las particularidades de las **herramientas OLAP**, es que sobre la base del modelado explicado son capaces de **agregar o preempaquetar todos las posibles peticiones de los usuarios previamente a las consultas**, de manera que los tiempos de respuesta del usuario son muy bajos, al estar ya todo o parte calculado. Esto es posible, ya que agregan los datos siguiendo la jerarquía de las dimensiones definidas.

II. DESCRIPCIÓN FUNCIONAL DEL SISTEMA

El sistema DATATUR consta de diferentes niveles de información. Su eje cen-

tral se basa en que los datos que proporciona para el grupo principal de usuarios, están modelados dimensionalmente, es decir, están almacenados en un sistema multidimensional. Esto implica muchas ventajas para este usuario final, ya que puede «navegar» por la información de una manera sencilla, sin tener que conocer ningún tipo de lenguaje de programación, exclusivamente mediante el uso del ratón, puede componerse sus informes a medida y hacer las selecciones que considere necesarias.

En contrapartida, el sistema está muy interrelacionado con el negocio, lo que requiere que los usuarios, por una parte, conozcan conceptualmente el modelo por el cual quieren navegar o extraer la información, y por otra, necesiten de unos conocimientos del negocio turístico o en concreto de la operación o área de estudio sobre la cual estén trabajando.

Los diferentes niveles de datos que contiene DATATUR se han dividido dentro del sistema en diferentes bases de datos: la de **Datos Primarios** o datos fuente de la operación, que en el caso de operaciones estadísticas en las que es necesaria la elevación de los datos, se corresponde con los ficheros pretabulables (a cada registro se le ha asignado un peso específico), la **Base de Datos Agregados** que se identifica con los modelos multidimensionales los cuales se explicarán ampliamente más adelante y la **Base de Datos de Metadatos** o de información sobre el dato estadístico almacenado en las otras dos, es decir, definiciones de variables, comentarios, documentos metodológicos asociados...

II.1. Base de datos primarios

El almacenamiento de los datos fuente se realiza en una base de datos relacional. Los datos originarios cargados proceden de fuente diversas, dependiendo no sólo del organismo elaborador, sino también del tipo de datos.

Existen dos tipos de datos fuente principalmente, los datos derivados de encuesta, que han sido elevados y tabulados para hacerlos corresponder a una población y que son principalmente los datos que explota el propio IET y los datos que se corresponden con directorios (ejemplo, hoteles, apartamentos) o datos estadísticos no elevados (ejemplo, Balanza de Pagos por Turismo). La diferencia fundamental entre unos y otros es el tratamiento posterior que ha de realizarse sobre ellos, antes de convertirlos en multidimensionales.

Dentro de la base de datos primarios se han creado tres grupos de tablas de datos diferenciados, tablas de datos de entrada, tablas de dimensiones y tablas de hechos.

Las tablas de datos de entrada corresponden como su nombre indica con los datos fuente, estos son transformados para poder verlos luego dimensionalmente y se convierten en tablas de hechos. Las tablas de dimensiones son creadas con los valores de los diferentes estados que pueden tener los elementos dentro de cada dimensión (véase apartado de conceptos).

Para ilustrarlo más claramente vamos a ver un ejemplo real que se corresponde

con una de las principales operaciones estadísticas del Instituto de Estudios Turísticos, Frontur (Movimientos Turísticos en Fronteras). Esta es una operación derivada de encuesta y los ficheros de entrada para DATATUR son del tipo «pretatabulables», cada registro del fichero tiene asignado un peso (representa a x individuos).

Dentro de esta operación existen diferentes suboperaciones, entradas por carretera, entradas por aeropuertos y entradas por tren y barco. Cada una de estas suboperaciones tiene diferentes ficheros de datos muestrales, por lo que existen diferentes tablas según sea la suboperación y el año de los datos. El modelo dimensional que se ha diseñado para Frontur Entradas es, sin embargo, único, ya que se pretendía unir toda esta información, por lo que en un primer paso del diseño hubo que seleccionar aquellos campos comunes en todas las suboperaciones y seleccionar los indicadores.

En algunas suboperaciones las transformaciones necesarias no fueron muchas, pero otras como en el caso de entradas por carretera se necesitaron muchos cambios. Entre los más significativos se pueden enumerar los siguientes:

- Uniones de campos: pernoctaciones extranjero + pernoctaciones en España a un único campo de pernoctaciones.
- Desestimación de campos: no tener en cuenta los campos de agrupaciones superiores de países, ya que las agregaciones están definidas en las tablas de dimensiones.

- Generación de campos que no existían: en los modelos se ha creado una dimensión que es la periodicidad del dato (mensual, acumulado) y se ha tenido que crear un campo que haga referencia a esta variable.
- Limpieza de algunos campos: comprobaciones sobre los estados de una variable para evitar incongruencias en los cruces (que un excursionista que por definición no pernocta no pueda tener asignado un tipo de alojamiento).
- Además, para unir las distintas suboperaciones hubo que revisar las codificaciones de los campos por los que se unieron las tablas (evitar que furgoneta en carretera y vuelo charter en avión tengan el mismo código).

El resultado al final del proceso fue una tabla de hechos con los datos de todas las

suboperaciones y de todos los periodos, a la que actualmente, a través de los procesos de actualización diseñados, se van añadiendo nuevos registros según van llegando más datos de nuevos meses.

A la tabla de hechos resultante de todo el proceso anterior se unieron por medio de las claves, las tablas de dimensiones para dar como resultado el esquema en estrella explicado en el apartado de conceptos. El resultado puede comprobarse en la figura 3.

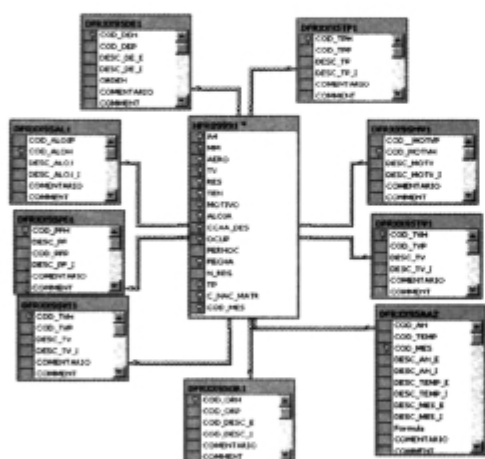
II.2. Base de datos de agregados

Las bases de datos agregados o multidimensionales, constituyen el núcleo del sistema, ya que es sobre ellas, sobre las que se realizan casi todas las consultas.

Estas bases de datos no son de tipo relacional y el software empleado para su desarrollo no es tan conocido como el de las anteriores, pero no es complicado. La mayor complejidad en este tipo de desarrollos es la parte de análisis, el diseño propio de la base de datos más que su desarrollo e implantación.

Las BD dimensionales permiten generalmente distintos tipos de almacenamiento conocidos como MOLAP, ROLAP y HOLAP (1). En nuestro caso, al ser las bases de datos no muy grandes (aunque contengan millones de registros) y siendo, sobretodo, lo que prima los tiempos de respuesta, hemos optado por el almacenamiento MOLAP, en el cual todos los datos, incluidos los de origen (existe una redundancia de datos), se almacenan en el formato multidimensional,

Figura 3



que es el almacenamiento que mejores tiempos de respuesta ofrece.











Para la operación estadística de Frontur Entradas, la base de datos dimensional o «cubo», consta de 10 dimensiones o puntos de vista sobre los que se puede analizar los indicadores de viajeros (personas que han pasado por la frontera) y pernoctaciones (véase figura 4).

Esto implica en la práctica que un usuario puede cruzar cualquier estado de las variables o dimensiones con cualquier otro. Para dar una mejor visión de lo que esto representa diremos que el número de combinaciones posibles que un usuario podría hacer sería el producto de los elementos o estados de todas las dimensiones que en el caso de la operación que estamos poniendo como ejemplo ascendería a 39.798.088.634.304. Esto es el número de

consultas teóricas que se pueden hacer sobre la base de datos, pero no en todos los cruces tendremos información, no todos los turistas provenientes de Suiza en coche y que hayan entrado por el punto fronterizo de «La Junquera» durante enero de 2000 han pernoctado en Segovia, si hiciésemos este cruce de variables seguramente no nos devolvería ningún dato.

La forma de efectuar la consulta anterior sobre el modelo sería a través del ratón, el usuario ha de desglosar la dimensión «País de residencia» y seleccionar Suiza, desglosar los periodos hasta llegar el elemento Enero 2000, desglosar puntos de entrada hasta «La Junquera» y seleccionar qué quiere ver como filas y columnas; por ejemplo, podría seleccionar como filas los motivos del viaje y como columnas los tipos de viajero (véase figura 5).

Figura 4

	ALOJAMIENTOS
	DESTINO PRINCIPAL
	MOTIVOS VIAJE
	PAIS RESIDENCIA
	PERNOCTACIONES
	TIPO TRANSPORTE
	TIPO VIAJERO
	VIAS DE ACCESO
	PERIODICIDAD
	PERIODO DE REFERENCIA

II.3. Base de datos de metadatos

Un complemento indispensable a la hora de dar información estadística son los «metadatos» considerando como tales a todo dato susceptible de aportar más información o claridad sobre el dato que se está consultando, o sea, cualificar el dato, ya sea por medio de definiciones de variables, documentos asociados, «links» o referencias a otros datos...

Los metadatos son esenciales para comprender correctamente los datos numéricos y para valorar la comparabilidad de distintas fuentes.

Como base de soporte de estos, se ha creado una base de datos relacional con esta

Figura 5



información que está estrechamente relacionada con las bases anteriormente descritas. Es muy importante que cuando uno está consultando un dato estadístico sepa, por ejemplo, la diferencia entre excursionista o turista o que meses componen los datos de una temporada o poderse leer el/los documento(s) de metodología asociados.

Una de las principales razones por las que se optó en su momento en hacer un desarrollo a medida en el IET, fue que ningún software actual, es capaz de mostrar información estadística pura con metadatos a través de Internet y esto es uno de los pilares fundamentales sobre los que se sustenta el sistema.

Dentro de esta base de datos, a través del programa especialmente diseñado para ello, los usuarios de mantenimiento del sistema son capaces de añadir definiciones o aclaraciones tanto a nivel dimensión, gru-

pos de elementos de una dimensión, información para distintos cruces de variables, incluso definiciones para los distintos estados de una variable o dimensión. Además se pueden asociar al sistema documentos en distintos formatos, relacionados con operaciones, suboperaciones, distintos meses, temporadas y años que son publicados automáticamente en Internet por el sistema.

III. ARQUITECTURA DE DESARROLLO

El sistema creado consta de los siguientes elementos que cubren todas las funcionalidades:

- Bases de datos relacionales y multidimensionales.
- Programas de extracción, limpieza y transformación de datos (ETL).

- Programas para la generación y publicación de información de las BD de manera automática en Internet.
- Interfaces de usuario para la consulta de datos vía web.
- Programas de usuario para la actualización y mantenimiento de los metadatos.
- Procesos de backup y seguridad del sistema.

Los programas **ETL** son un elemento imprescindible del sistema para su permanencia en el tiempo. Se encargan de facilitar al usuario técnico el mantenimiento y actualización de la información al provenir de diferentes fuentes. Son procesos automatizados hasta el nivel más básico posible (sólo necesitan del usuario que les informe de la fecha de los datos a cargar). La generación de estos paquetes de transformaciones de datos ha sido compleja debido a la diversidad de formatos y ha representado un tiempo importante dentro del desarrollo.

Otro elemento importante ha sido conseguir la publicación automática de distintos informes en Internet. Se ha creado un programa que en función de unos parámetros almacenados en la base de datos de metadatos permite al Instituto de Estudios Turísticos generar los informes mensuales que necesita publicar en Internet, ya sea para los usuarios públicos como para distintos organismos como Comunidades Autónomas u Oficinas Españolas de Turismo en el Extranjero (OET). Se pueden publicar informes mensuales, anuales y de temporada

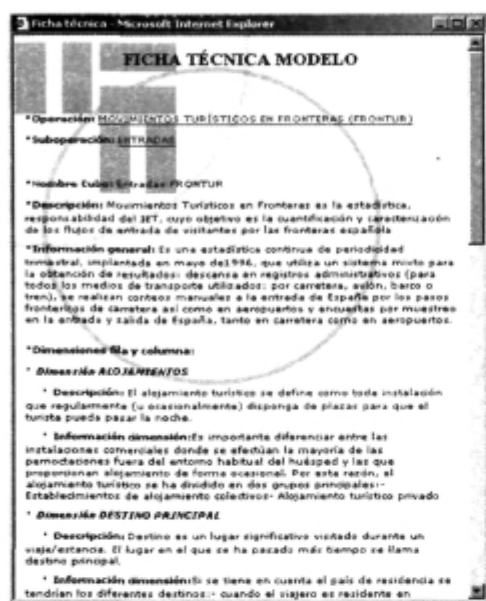
y el programa se encarga de acceder a las bases de datos para extraer la información en función de la parametrización, buscar sus metadatos asociados y generar todo en formato HTML (incluyendo generación automática de gráficos en función de los datos de los informes). Esto supone un gran avance, ya que esta publicación antes era manual y sólo se podían publicar pequeñas cantidades de datos, ahora se están generando en torno a 4.000 informes, con el consiguiente ahorro de tiempo y, por tanto, de costes.

La parte más visual del sistema y por la que la mayoría de los usuarios trabajan es la interfase vía web. En ella se ha conseguido dar al usuario una herramienta de trabajo intuitiva y potente, para trabajar a través de Internet, sin descuidar el diseño. Es la parte que junto con el tratamiento de los datos más esfuerzo ha requerido, pero los resultados lo han merecido.

La interfaz consta de dos partes, una, es donde el usuario se compone a través de movimientos de ratón y navegación «drill-up» y «drill-down» el informe que necesita, y otra, es la ficha técnica que nos da información sobre los datos que estamos viendo (metadatos). Para no sobrecargar el sistema, la ficha técnica sólo se actualiza cuando lo solicita el usuario y no cada vez que se cambian los datos en pantalla (véanse figuras 5 y 6).

El programa de actualización de metadatos permite a los usuarios técnicos añadir, borrar o modificar registros en la base de datos de metadatos sin que sea necesario que conozcan su estructura. Además, se ha

Figura 6



creado otra aplicación para gestionar los documentos o ficheros (Word, Excel, etc.), con los que trabaja el personal del IET, con la posibilidad, si lo desea el usuario, de que su documento se publique automáticamente en Internet asociado a una operación, suboperación y periodo.

Por último, se han definido los procesos automáticos de salvaguarda de toda la información de los diferentes tipos de bases de datos para su ejecución desatendida.

IV. ENTORNO TECNOLÓGICO

Todo el desarrollo realizado habría sido prácticamente imposible y por supuesto los resultados no habrían sido los mismos si el proyecto se hubiese realizado tan sólo hace unos años. Los avances que se producen en

el software y herramientas que existen en el mercado son muy rápidos y lo que ayer era una ficción, hoy es una realidad. Este ritmo de cambio, además, se va a mantener o aumentar en los próximos años.

Por esta razón, los productos con los que se inició el proyecto y con los que se ha finalizado, no son los mismos, en cuanto a la versión y, por tanto, tampoco las funcionalidades que soportan.

La plataforma utilizada ha sido la de Microsoft y las herramientas con las que se ha finalizado el proyecto son:

- En servidor:
 - Windows 2000 Advanced Server.
 - Microsoft SQL 2000.
 - Microsoft Analysis Services 2000.
 - Internet Information Server 5.0 (IIS).
 - Microsoft Component Services.
- En los puestos de desarrollo:
 - Windows 2000 Professional.
 - Microsoft Visual Interdev 6.0.
 - Microsoft Visual Basic 6.0.
 - Front Page 2000.
 - Microsoft Office 2000.
 - Cliente de SQL Server 2000.
 - Cliente de Analisis Services.

El proyecto comenzó con los mismos productos pero en sus versiones anteriores (1) (Microsoft SQL 7.0, Windows NT 4.0, OLAP Services 7.0), pero hubo que migrar principalmente por la necesidad de utilizar las nuevas funcionalidades de mo-

delado multidimensional de la nueva versión de los servicios OLAP y SQL 2000 y por necesitar comprimir la información a través de Internet (debido al tamaño de algunos informes), característica soportada por el IIS de Windows 2000.

Todo el proyecto, como se ha ido destacando a lo largo del artículo, está orientado a la difusión de datos a través de Internet, por tanto una gran parte del esfuerzo se ha centrado en este punto. Todos los desarrollos se han realizado pensando en que no fuese necesario instalar nada en los puestos cliente, ni siquiera componentes Java ni controles OCX. Para ello se han programado las interfaces con Microsoft Visual Interdev utilizando Active Server Pages (ASP), que es una programación que ejecuta el servidor para devolver páginas distintas según las peticiones del usuario. En el lado de cliente se ha utilizado Java Script para hacer las funciones necesarias.

Las aplicaciones de este tipo son muy sensibles al número de usuarios concurrentes que hacen peticiones de datos desde Internet, por lo que para evitar saturaciones futuras y que el sistema sea escalable, se ha creado también mucha parte de la programación del servidor en componentes, que utilizando los servicios de Microsoft Component Services hace que la escalabilidad sea sencilla.

El sistema ha sido desarrollado para el soporte de diferentes idiomas en el futuro, por lo que se han creado distintos componentes para que se comporten como si de usuarios con distintos idiomas se trataran. Esto hace posible desde Internet la selec-

ción del idioma (según el idioma elegido se utilizará un componente u otro para conectarse a las BD). En principio está previsto el español y el inglés, pero se pueden añadir los que se quieran, el trabajo por supuesto está en la traducción de los estados de las variables y de los metadatos.

V. TRATAMIENTO DE LA INFORMACIÓN

Un sistema de información turística, Statistical Information System (SIS), realiza una serie de procesos estadísticos y maneja unas categorías de datos, datos estadísticos.

En una organización estadística existen procesos para los siguientes tipos de tareas:

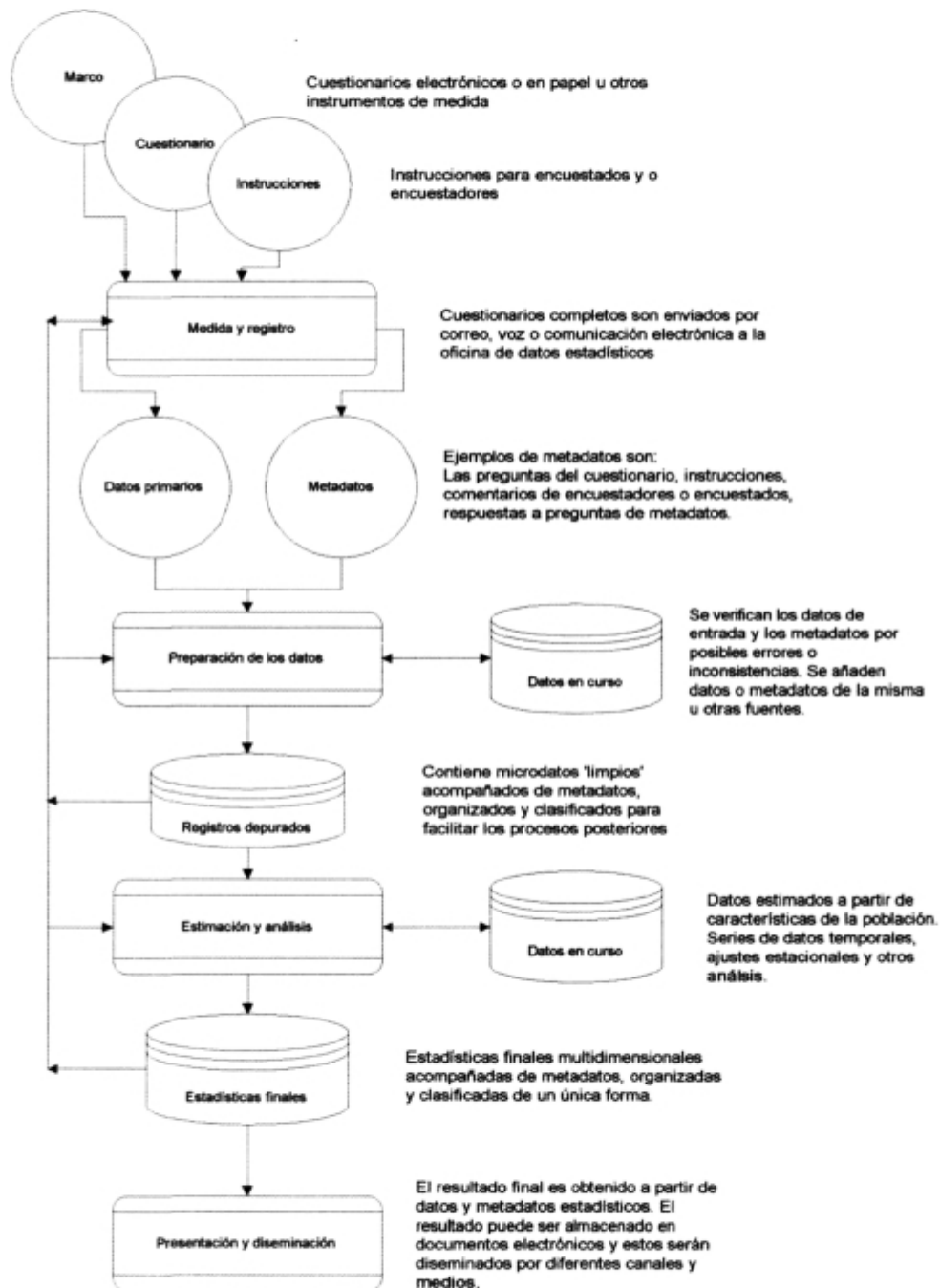
- Recogida de datos.
- Elaboración de datos.
- Almacenamiento de datos.
- Recuperación de datos.
- Análisis estadístico de datos.
- Difusión de información estadística.

Los procesos estadísticos usan y producen datos estadísticos, estos según definición de la Comisión Estadística de la OCDE, pueden ser microdatos, macrodatos y metadatos. Cada uno de ellos ya han sido definidos en este artículo.

El sistema DATATUR se ha desarrollado para facilitar y agilizar los cuatro últimos procesos, el almacenamiento, recuperación, análisis y difusión.

Aunque el sistema cubre gran parte de los procesos de una organización dedicada a

Figura 7
Flujo de datos y metadatos típico a partir de una encuesta estadística



Fuente: Elaboración propia a partir de los trabajos de la Comisión Estadística y Económica para Europa. Work session on Statistical Metadata (METIS) Berlín, Alemania, 22-25 octubre 1996.

elaborar datos estadísticos, también es cierto que sólo es la punta del iceberg de un trabajo ingente que hay en los procesos previos que son la recogida y elaboración de los datos.

Hasta que los datos ya elaborados llegan a DATATUR se han tenido que realizar en el caso de datos provenientes de encuestas, la definición de estas, la toma de datos en campo, el estudio de las tomas para la mejora de la recogida, las limpiezas previas de las encuestas, la tabulación y elevación de los datos, etc.

Todas estas tareas conllevan el uso de otras herramientas más específicas adaptadas a cada una, como es el uso de PDA (Personal Digital Assistant) en la recogida de datos, el uso de software especializado en la elevación y depuración de datos.

Por tanto, DATATUR se nutre de unos procesos y tareas previas sin los cuales no podría existir. El esquema completo del flujo de datos y metadatos de una operación estadística, a partir de una encuesta se puede ver en la figura 7.

VI. PERFILES DE ACCESO AL SISTEMA

Se han creado distintos perfiles de acceso al sistema atendiendo sobre todo al conocimiento no tanto de la herramienta sino de las operaciones. Como hemos hecho referencia en otros puntos de este artículo el acceso a los modelos de datos dimensionales no necesita de conocimientos informáticos de ningún tipo, pero sí del modelo o el negocio.

Al accederse a los datos con tanta facilidad es sencillo que un usuario saque datos sin fijarse realmente o sin conocer lo que está viendo. Es fácil tomar unos datos como si fuesen turistas cuando realmente lo que estamos viendo son viajeros (turistas y excursionistas) o incluso no conocer la diferencia.

Por ello hemos creado tres perfiles de acceso a los datos:

1. El usuario público.
2. El usuario de suscripción con acceso restringido.
3. El usuario de suscripción con acceso completo.

El usuario público sólo accede a la información a través de informes predefinidos que ya han sido generados para él y puestos a su disposición a través de Internet. No puede componerse sus informes, pero a cambio tiene la seguridad de que la información es correcta. Este usuario puede exportar los datos de los informes predefinidos o imprimirlos. A este grupo de usuarios pertenece el público en general. Este perfil cubre las necesidades de datos de la mayoría de los usuarios potenciales.

Al usuario con acceso por suscripción, pero con acceso restringido, se le deja hacer sus análisis y componerse sus informes porque ya tiene unos conocimientos suficientes para ello, pero no se le deja bajar hasta los datos elementales de la operación, donde la muestra es insuficiente para que los datos que se cruzan sean significativos.

El último tipo de usuario es el que puede acceder completamente a los datos del sistema.

Los dos usuarios por suscripción deben cumplir con unos requisitos de conocimientos y responsabilidad en el tratamiento de la información que se les proporciona con la herramienta. Antes de la publicación de resultados extraídos de los datos del sistema deben tener el consentimiento del Instituto de Estudios Turísticos y deben citar la fuente en todos sus estudios.

VII. PERSPECTIVAS DE FUTURO

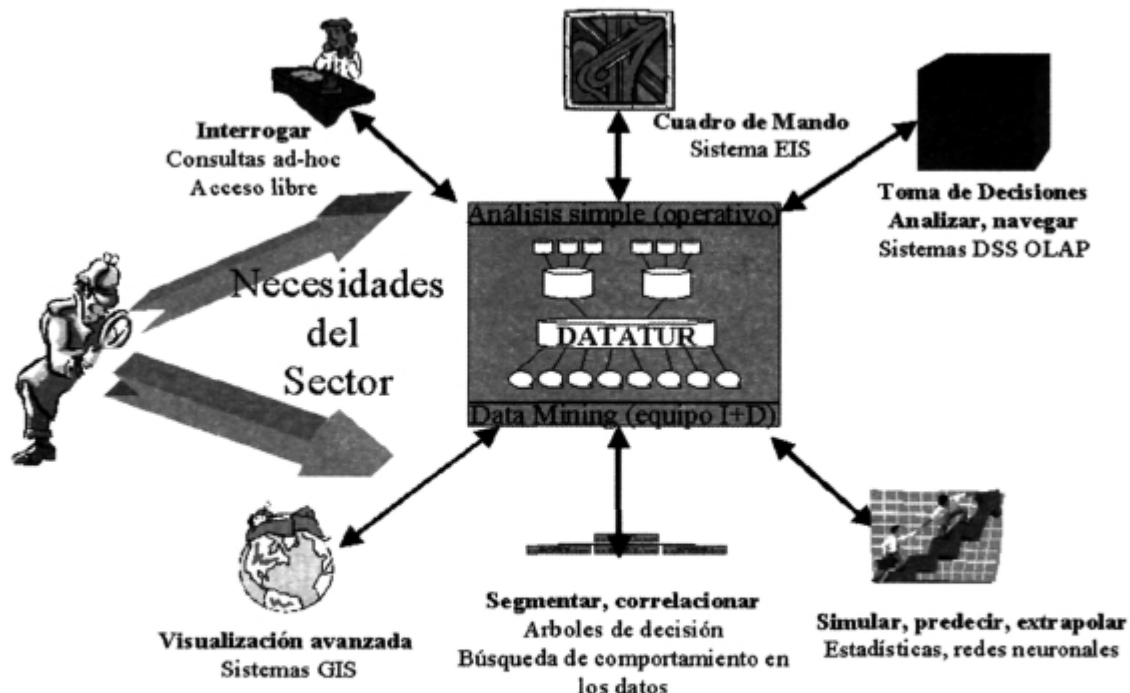
Casi es obligado concluir como comenzábamos el artículo, haciendo hincapié en

que un Data Warehouse, como es el sistema que hoy nos ocupa, debe evolucionar para hacer frente a los cambios que la realidad turística requiere o caer en el olvido y el desuso. Por tanto, es obligado que se amplíe no sólo en la información que contiene, llegando a acuerdos con otras instituciones turísticas para incorporar sus datos, sino también, en sus funcionalidades y modos de acceso.

DATATUR al ser un almacén de datos estructurados y clasificados en formato electrónico es potencialmente accesible de múltiples maneras.

Hasta este momento se han desarrollado o utilizado para su consulta herramientas de análisis OLAP, consultas predefinidas

Figura 8



para usuarios de acceso libre y herramientas de hoja de cálculo para el trabajo de los usuarios internos del IET, donde además de analizar pueden representar la información de manera gráfica. Pero estas no son las únicas maneras por las que se podría tratar la información.

Todos los tipos de acceso que en un futuro se podrían utilizar están representados en la figura 8.

Como se puede ver, y a modo de conclusión, todavía quedan muchos caminos por los que mejorar y potenciar DATATUR, desde visualización cartográfica de la información (Sistemas GIS), pasando por herramientas de Data Mining y cuadros de mando, para recoger de una manera organizada y orientada a dar una visión global y rápida, las últimas tendencias del turismo en cuanto a información estadística.

BIBLIOGRAFÍA

FRANCO, JEAN MICHEL y EDS-INSTITUT PROMÉTÉUS (1997). *Le Data Warehouse, le Data Mining*. Editions Eyrolles, Paris; Ediciones Gestión 2000, S. A., Barcelona.

INMON, W. H. (1996). *Building Data Warehouse* (2.ª edición). John Willey & Sons (Sd).

INMON, W. H. y HACJATHORM, RICHARD (1994). *Using the Data Warehouse*. Q E D Pub co.

INSTITUTO DE ESTUDIOS TURISTICOS (2000). *Estadística de Movimientos Turísticos en Fronteras (Frontur)*. Referencias metodológicas. Madrid. Instituto de Estudios Turísticos. Secretaría General de Turismo. Ministerio de Economía.

INSTITUTO DE ESTUDIOS TURISTICOS (2000). *Estadística de Movimientos Turísticos de los Españoles (Familitur)*. Referencias metodológicas 2000-2001. Madrid. Instituto de Estudios Turísticos. Secretaría General de Turismo. Ministerio de Economía.

KIMBALL, RALPH (2000). *The Data Warehouse Toolkit*. John Willey & Sons (Sd).

MICROSOFT (1999). *Estrategia de Data Warehousing de Microsoft: una plataforma para mejorar la toma de decisiones mediante un acceso a datos y un análisis más fácil*.

NOTAS

(1) INSTITUTO DE ESTUDIOS TURISTICOS (1999): «DATATUR. Sistema de Información de Estadísticas Turísticas», *Estudios Turísticos*, núm. 142, pp. 37-56. Madrid, Secretaría de Estado de Comercio, Turismo y Pymes.