Spatial Data Warehousing for Hospital Organizations

An ESRI White Paper—March 1998

Spatial Data Warehousing for Hospital Organizations

An ESRI White Paper

Contents **Page**

Spatial Data Warehousing for Hospital Organizations

Data Warehousing

Data warehousing is a $15 billion global market, growing 10 to 15 percent annually. It is estimated that 95 percent of Fortune 1000 companies and many government and service organizations are implementing data warehouses to make sense of the vast amount of information they need to operate and plan. What is this phenomenon that is changing the way business captures and analyzes data?

A data warehouse has been defined as a subject-oriented, integrated, time variant, nonvolatile collection of data that support a company's decision making process. In essence, a data warehouse is a large database organizing operational data in a repository for easy query and analysis. It is a well conceived and well designed environment containing data that are key to an organization's decision making process. By definition, a data warehouse is a large database designed to support the decision making needs of an organization.

Spatial Data Warehousing

Data warehousing applications are based on high-performance databases that use a client/server architecture to integrate diverse data types in near real time. While data warehouses look at many types and dimensions of data, many are lacking in the spatial, or location, context of the data. Over 80 percent of health care data have some spatial context such as a patient's address, ZIP Code, or provider location. By using technology that integrates this spatial component with the data warehouse, an organization can unlock this hidden potential in their data, allowing them to see hidden relationships and patterns in data, in essence data mining by geography.

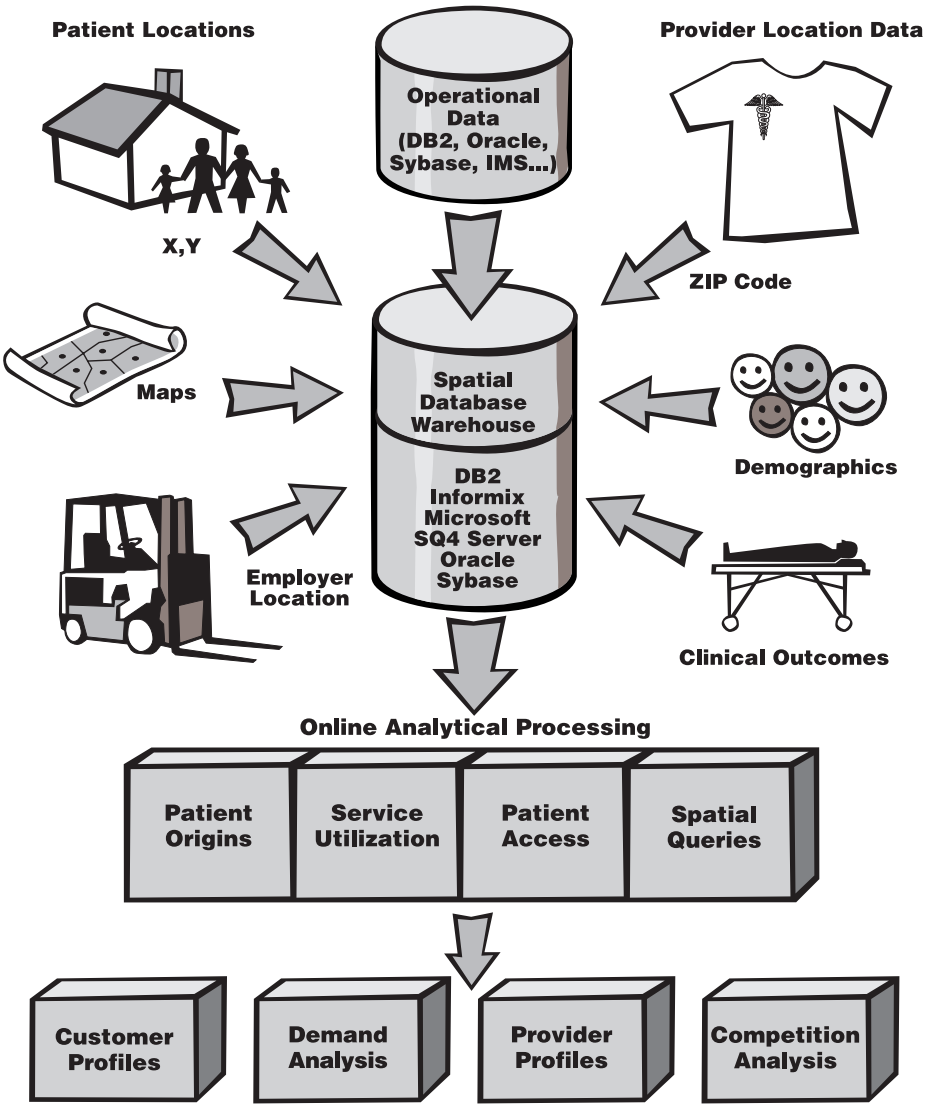
This white paper will explain how spatially enabling your database will benefit your organization with

- More organized data structure
- Better integration of disparate data
- New, spatially enabled analysis
- Reduced decision cycle time
- Improved decisions

The Need for a Spatial Data Warehouse

Why do you need a spatial data warehouse? Most hospitals have built elaborate operational systems for collecting data. Operational data in a hospital are data that allow it to function each day. These operational data include data from patient registration, medical records, billing, laboratory, and departmental systems, which are designed to allow managers to run a department, not analyze it. Operational systems data tend to be process-oriented and narrowly focused. In a hospital, for example, operational data include items such as patient demographics and clinical data captured in various

departments. Patient registration systems capture locational data by street address and ZIP Code. To be analyzed, these data must be integrated with standard attribute data into an enterprisewide repository from which decision makers can perform ad hoc analysis and run reports. A spatial data warehouse enhances the data by storing and analyzing the spatial component of operational data. This new dimension, location, gives decision makers more definition of their data and allows them to ask new questions about relationships in their database.



Online Analytical Processing

The spatial data warehouse extends the usefulness of online analytical processing (OLAP) systems. OLAP systems are used by decision makers to interrogate the data warehouse. The interrogation is supported by OLAP, sometimes also referred to as

multidimensional online analytical processing (MD-OLAP) or relational online analytical processing (ROLAP). The biggest difference between ROLAP and OLAP is how the query is handled by the database system. In OLAP, data are aggregated into a separate database. With ROLAP, requests are made directly to the database.

The data for analysis with OLAP are accessed through metadata that document data source, frequency of update, and location of data. The data returned from the queries are represented as "multidimensional," although their form may be maintained as relational. A multidimensional database is a database where data are structured as measures and dimensions. Measures are numerical data such as patient volume by diagnosis, procedure, financial class, or physician. Dimensions are the kinds of data that can be summarized with measures such as physician location, service area, referral region, or distance. The user can specify high- or detailed-level views of data with database navigation through drill downs in reports to finer levels of detail and analysis by product line, geographic location, diagnosis, or payor.

Drill downs allow a user to ask more detailed questions based on the answers they have received from OLAP or ROLAP. For example, after identifying the most profitable diagnostic related groups (DRGs), the user can then search the spatial data warehouse for information about the demographic characteristics of the patients who were assigned that DRG, the various diagnoses associated with these DRGs, and which physicians generate these cases.

Thus, a data warehouse that includes a spatial data model adds new analytical functions to OLAP including patient profiling and physician profiling.

In order for the data warehouse and OLAP to function properly, the software for the data warehouse and applications must be open, so when new components are added to the collection of the data, they can be integrated into the system. The software should also be scalable to accommodate increasing use, different types of usage, and increasing amounts of data.

ESRI's Solution to the Spatial Data Warehouse

ESRI is the largest provider of spatial technology worldwide. This spatial technology, known as a geographic information system (GIS), allows users to view, query, and analyze their business data based on the locational context of their data. ESRI's family of software products meets the need to spatially enable your data warehouse. ESRI's GIS software is geared toward specific user needs unified by a common approach to working with spatial information. Data created and maintained with one GIS product can be analyzed and displayed using another. You can create your own GIS configuration by selecting appropriate solutions from ESRI's comprehensive family of products as your GIS requirements change or grow. ESRI® software products can easily be deployed across enterprisewide installations, allowing each user to choose the tools they need. ESRI distinguishes itself from competitors with the breadth and scalability of its software solutions offerings. ESRI is unique from any other vendor in this regard.

Spatial Database Engine

ESRI's Spatial Database Engine™ (SDE™) software is the core component in ESRI's solution for data warehousing. SDE is a high-performance, object-based spatial data access engine using open standards and a client/server architecture. SDE provides

powerful tools that can be used for simple or complex data analysis on very large spatial databases. SDE stores spatial data in commercial database management systems (DBMSs). The spatial data are in relational tables and allow users to perform indexing, compression, backup, and security.

ESRI is working closely with IBM, Informix, Microsoft, Oracle, and Sybase to port SDE to the latest releases of their software solutions.

SDE is highly customizable, so organizations can develop focused and specific applications that provide access and manipulation capabilities for virtually any type of spatial data.

SDE in the Data Warehouse

SDE provides both a data model to the data warehouse and a geographic analysis engine for OLAP. SDE allows users to store spatial data inside the data warehouse. SDE offers data transformation and manipulation, a spatial storage engine, robust data access mechanisms, and a broad range of analytical tools and methods that are designed to facilitate spatial analysis and are open to incorporating other vendors' products. Therefore, SDE performs geographic analysis and query better than DBMSs that focus on online transaction processing.

To understand how SDE and ESRI's products enhance the data warehouse, we will revisit the data warehouse definition.

First, SDE supports the data warehouse by being a component of an organization *standard* database such as IBM, INFORMIX, Oracle, and SYBASE.

Second, the data warehouse must be *subject oriented*. This kind of data includes spatial-based data that are contained in point-of-service information, home addresses, employer address, and service location.

Third, the data must be *integrated*. Because SDE stores all geographic data and their attribute data in tables in an DBMS, the spatial data and their attributes can be related or joined to any other table in the DBMS.

Fourth, the data must be *time variant.* Spatial data can be "time stamped" at the level of individual or grouped spatial features. Spatial features can also be time stamped by maintaining different versions of features or groups of features. In the latter case, the location of features can be shown at time x, then the same set at time y.

Finally, data should be *nonvolatile*. SDE comes with a number of clients that can perform editing on spatial features.

Thus, by being a technology component of standard databases, SDE is ideally suited to be at the center of ESRI's data warehousing solution.

SDE Data Warehouse Architecture

To describe how SDE fits into the data warehouse, the functional model of a data warehouse will be followed. The functional model describes the most important components of the data warehouse from data loading to end user application development.

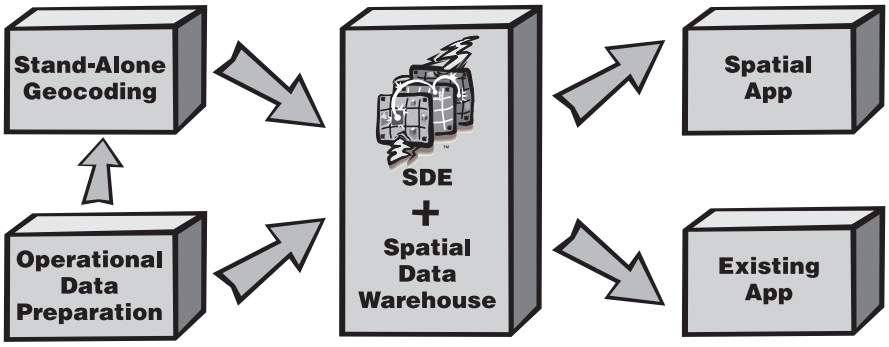
Data Loading Through Standard Published Data Formats

SDE stores all spatial features in one common data model and can convert data from many sources into that data model. The most commonly used format for loading data into SDE is the shapefile. The SDE data model standardizes the input of spatial data into the spatial data warehouse through ESRI's shapefile format. The shapefile is an open, published format from ESRI that is supported by all ESRI products. The published shapefile format is readily accessible by business users and application developers and can be easily supported in their systems.

Spatial Data Preparation

Consistency and integrity checks are the same for any spatial feature regardless of whether the feature is a point (such as an address), a line (such as a street), or an area (such as a ZIP Code boundary). These checks are applied before the data are entered into the data warehouse. For example, addresses can be matched and associated with a location through software from ESRI such as ArcView® GIS. These addresses can then be used to validate data from operational or external sources before they enter the data warehouse. Reports showing the success of address matching are produced as part of the matching process and erroneous or missing patient address information can be identified before its inclusion into the data warehouse. Displays created by the spatially represented data can show areas of omission or excess coverage of areas that are of limited interest. These displays can be used to cull data sets before they enter the data warehouse.

Spatial Data Warehouse Flow



Spatial Data Warehouse Data Model

As the spatial enabling technology for data warehousing, SDE offers a continuous, nontiled data model. Key benefits of the SDE approach include the following:

- One integrated data model for points, lines, and areas
- Easy and efficient queries and integration with other types of data

Spatial objects managed by SDE are stored in a table with one row for each spatial object. The spatial objects can include points for data such as addresses, lines for streets, and areas for ZIP Code boundaries. The spatial objects are combined into seamless sets of data called layers. The layers can be attributed by scale for display so that the relationship of other data in the data warehouse to location can be established.

Spatial Data Analysis Tools

In the data warehouse, OLAP or ROLAP systems are the data analysis tools. These tools access and process data in the data warehouse. In some cases, multidimensional views are created, and in other cases the data are processed for output to business applications.

OLAP or ROLAP systems can use SDE software's open C callable application programming interface (API) consisting of over 120 functions. These functions include spatial search and spatial analysis. For example, the tools can be used to select patients or providers within a ZIP Code area or distance.

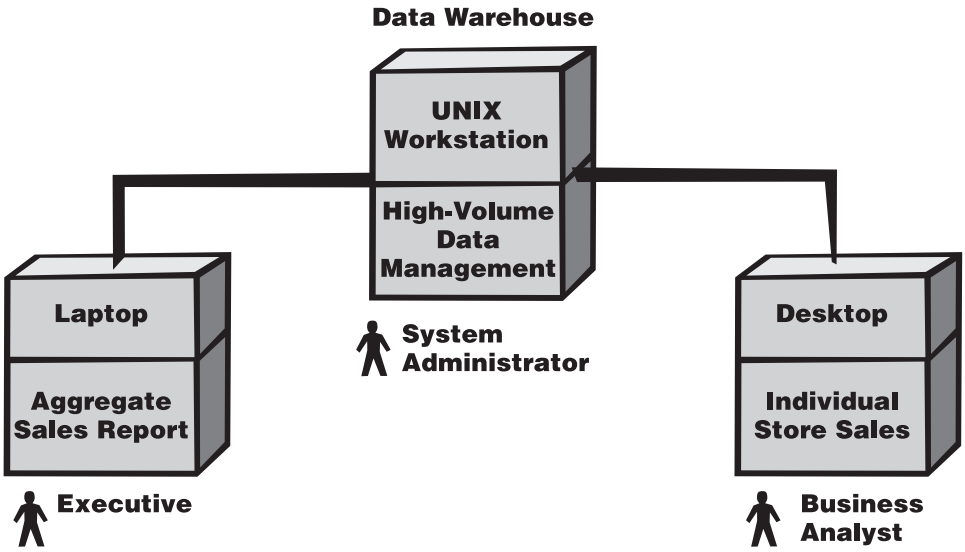
ESRI also provides off-the-shelf GIS clients that utilize the spatial data warehouse. These ESRI solutions, available separately, embed the SDE C API in MapObjects™ software, ArcView GIS, and ARC/INFO. Any of these tools will allow processing or filtering data geographically. Although the clients vary from "thin," single-purpose ActiveX® technologies to multifunctioning spatial analysis clients, ultimately each client uses the API functions of SDE.

Open and Scalable Architecture

In order for the data warehouse and OLAP to function properly, the software for the data warehouse and applications must be open and scalable. The data warehouse must be open so that new components can be added and integrated into the system. The data warehouse must be scalable to accommodate increasing use and volume of data. ESRI provides solutions on a wide range of development environments and platforms. For example, an executive with a laptop or portable PC viewing an aggregate sales report, a business analyst with a desktop unit looking at individual store sales, and an information technology (IT) professional with a UNIX® workstation administering or viewing raw data can all access the same spatial data warehouse managed by SDE.

J-7775

Open and Scalable Data Warehouse



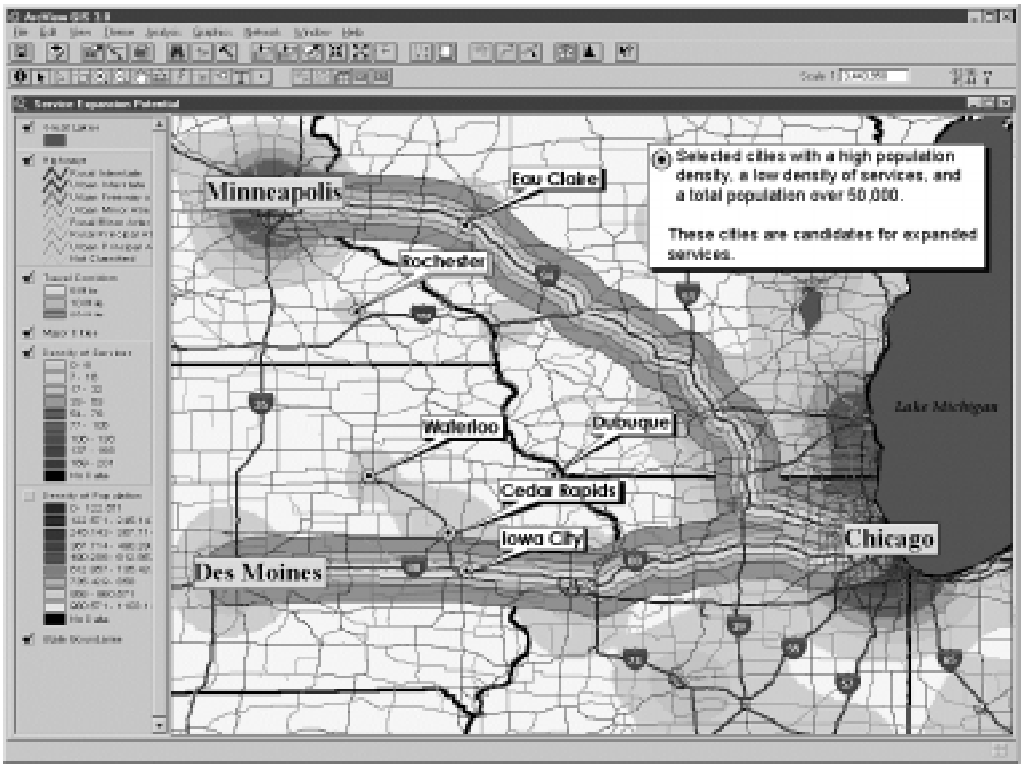
SDE Clients ESRI provides SDE client software products to meet organizations' data warehousing and OLAP needs: ArcView GIS, MapObjects, ARC/INFO, ArcExplorer, Internet clients, CAD clients, and the SDE C language API.

ArcView GIS ESRI's ArcView GIS is the world's leading desktop GIS product due to its powerful GIS functions and intuitive approach to analysis and data visualization. ArcView GIS on UNIX and Windows computers can be used out of the box as a client query tool for SDE. The ArcView GIS scripting language Avenue™ software and the ArcView GIS graphical user interface both support access to SDE. ArcView GIS is easily customized for SDE applications using Avenue software.

MapObjects ESRI's MapObjects can be used to embed mapping and GIS functionality in your existing IT application environment. ESRI's MapObjects is a collection of powerful mapping and GIS components that can be used to build SDE client applications in standard Windows® development environments such as Visual Basic®, Visual C ++®, and PowerBuilder®. MapObjects provides Windows developers with ActiveX-compliant access to SDE databases.

ARC/INFO ESRI's flagship software, ARC/INFO, gives you hundreds of built-in functions for sharing and processing geographic data, plus optional, fully integrated extensions for performing specialized analysis tasks. Running on UNIX or Windows NT®, everyone in the organization can access GIS data and perform sophisticated, high-end, custom geoprocessing operations with ARC/INFO. As a client application to SDE, ARC/INFO is designed for SDE data conversion, preparation, and loading; sophisticated cartographic production; and high-end spatial analysis.

ArcView GIS as a Client to SDE



ArcExplorer ESRI's lightweight GIS data explorer, ArcExplorer works with local data sets, performs as an SDE client, and operates as an Internet client to data and Internet map servers. ArcExplorer provides effective display and query tools as well as data retrieval capabilities. ArcExplorer is built using MapObjects and is representative of the type of SDE client that can be built using ESRI's development tools.

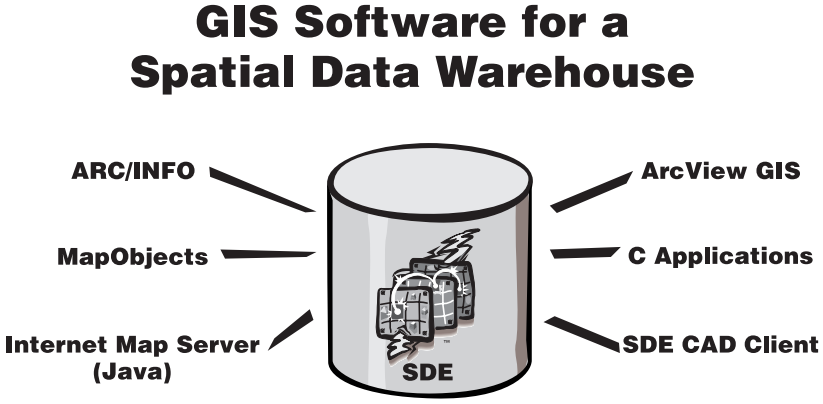
Internet Clients ESRI's Internet Map Servers are OLAP applications accessible via standard Web browsers. You can embed mapping in Web-based OLAP applications with MapObjects Internet Map Server or use ArcView Internet Map Server as an out-of-the-box query and analysis solution. ESRI's Internet Map Servers work with HTML and Java™, so browser plug-ins are not required. With ESRI's Internet Map Server solutions, newcomers to your market can access your medical staff services without specialized software.

CAD Clients SDE CAD Client software is an extension for MicroStation® and AutoCAD® software that allows CAD users to share and participate with the rest of the organization. With SDE CAD Client installed, these popular CAD packages gain access to SDE data. CAD objects stored with SDE CAD Client maintain their geometric definition, properties, and database connections. During the storage process, the objects are stored as pure CAD objects and also stored as SDE shapes. When a CAD client (such as MicroStation) performs a query, the selected CAD objects are returned. When a non-CAD client (such as ArcView GIS) performs a query, the SDE shapes are returned. The benefits are

twofold—CAD users get a logically continuous database and other users benefit from access to CAD information. Facility managers can realize tremendous benefit from this extension in their facility master planning.

C Clients

SDE includes a C language API on a variety of UNIX platforms as well as 32-bit Windows platforms (i.e., Windows 95® and Windows NT). The SDE client API provides the C and C++ programmer with complete access to SDE functionality.



Linkages between features and attributes can be established through ArcView GIS. Additional tools available through ARC/INFO and ArcView GIS even allow routing and spatial modeling to be applied to data in the data warehouse directly. Because SDE uses a client/server architecture where clients perform geoprocessing and the server performs retrieval of the data for geoprocessing, SDE applications are scalable and can be run on many types of clients. Output of OLAP to business applications is also supported on the desktop through the use of OLE and DDE.

A business analyst can easily view data to perform these tasks on a map by embedding a MapObjects application in an existing OLAP tool or by using ESRI's out-of-the-box desktop product, ArcView GIS. Both products support standard desktop formats and procedures, such as visualization and integration with spreadsheets, and have a relatively short learning curve. SDE also offers spatial search functions and can be integrated with OLAP tools. Depending on the requirements of the analyst, MapObjects, ArcView GIS, or SDE will perform these tasks on the data warehouse.

An Integrated Solution

From data loading to end user applications, SDE and ESRI's client products are important components of a successful data warehouse solution. Some of the benefits of using ESRI's solutions to data warehousing include

- Integration with "corporate" tabular data
- Stores a continuous, very large distributed database
- Very high performance
- Open APIs for flexible application development
- Multiuser access through client/server architecture

A substantial investment has already been made by the global community in geospatial and GIS data that already exist in ESRI format. Public agencies and private companies would want to exploit this easily available data in the data warehousing environment.

Business Applications

End user business applications are based on aggregating data and refining queries. For example, when monthly patient admissions or outpatient encounters are aggregated by service region and reports generated, individual physician contributions can be analyzed geographically. The aggregation is the dimensioning process described earlier, but because the geographic data are managed by SDE, the dimensioning can also be done spatially as well as through attribute fields with region names or numbers. Refinement of queries on the various service areas may produce physician or managed care figures and financial projections.

Data on individual providers can be used to correlate with information about patients taken from point-of-service data. Refinement of queries using geography can produce mapped or mapless output of demographic characteristics of patients by ZIP Code, drive time, market areas, or any other relevant geographic definition served by the hospitals in the region. The database can be queried to find areas of similar demographic content or geographic location relative to best practices and outcomes. Used in this way, the spatial data warehouse can help a hospital organization understand its various markets more precisely.

Consider this scenario: A product line manager is searching for the highest demand in a market area for a specific health "product." Having found the area, the manager then queries to find what are the unique geodemographic characteristics of their patients who have recently utilized the service. After the identification is made, the manager may want to profile this utilization and compare their utilization profile to the total patient populations residing within the hospital's primary marketing area. Using this information the manager may decide to replicate or adjust a practice development plan. Aggregated with other service location and patient data, "customer" product profiles may be created to craft marketwide advertising and promotion plans. This micromarketing analysis includes the use of both internal and external geobased data.

Benefits of the Spatial Data Warehouse

What are my natural service areas?

Where is my medical staff located in relationship to where patients reside?

Which employers have employees living closest to my primary care providers?

How effective was our recent advertising campaign in reaching our target market?

Patient and provider demographic data coupled with powerful geographic analytical tools can greatly increase the value of internally generated analysis. Here are a few ways that a spatial data warehouse can achieve significant results in a hospital organization.

- **Product Line Planning—**By linking operational data to patient and provider location, a product line manager can begin to visualize the geographical drivers of the product line. Issues such as the adequacy of patient access, and market penetration can be quickly evaluated and acted upon.

J-7775

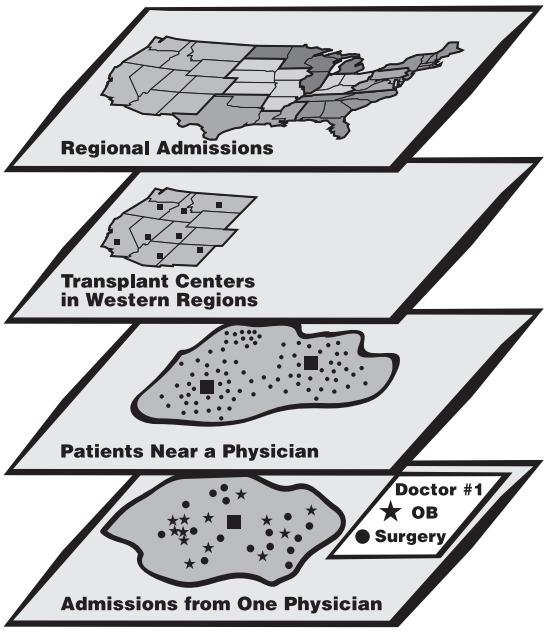
Benefits of the Spatial Data Warehouse

Where are my service areas? How should I define them?

Where are the physicians in my region? How close are they to each other?

What are the origins of my patients? Where do they live?

How effective was our recent marketing promotion? Where did our response come from?



- **Service Scheduling—**Health services delivered in the patients home can be scheduled in a much more efficient manner by analyzing transportation factors and street patterns and recommending the most efficient route to and between each patients home. The organizational efficiency of health service worker deployment can be easily accommodated within the analytical framework of the spatial data warehouse.

- **Marketing—**Getting the right message to the right market requires a fairly high degree of precision. Patient data which has been spatially enabled can be used to profile responses to various advertising programs and help refocus messages to targeted segments of the market.

- **Managed Care Contracting—**The profitability of most managed care contracts hinge on a combination of volume and price. A spatial data warehouse of service utilization information can be used to determine which plans, contracts, and/or networks provide the most cost-effective populations, while inappropriate utilization can be contrasted against a background of relevant population demographics.

- **Site Selection—**The time it takes to locate a new service or evaluate the acquisition of an existing service location can be significantly reduced by the geographical analysis of existing business in light of future demand. Presenting a picture of actual patient origins can enhance knowledge about new market opportunities.

- **Competitive Tracking—**The spatial data warehouse can significantly reduce the problem of keeping your competitor's service locations current. Maintaining accurate geographical information on competitors can be vastly improved by the use of a spatial data warehouse. The spatial data warehouse can provide the means by which every provider location, in the market is always ready to be illustrated correctly, geographically.

Industry Uses for Spatial Data Warehouses

Spatially enabled end user applications include the following categories:

- Strategic Planning
- Marketing Management
- Service Demand Forecasting
- Network Optimization
- Regulatory Compliance
- Telemarketing
- Customer Service
- Physician Relocation
- Database Marketing
- Facility and Property Management
- Target Marketing
- Competitive Analysis
- Direct Mail
- Market Research
- Product Line Marketing
- Advertising Validation
- Demographic Analysis

Trends in Data Warehouse Publishing

Among other trends in data warehousing is data publishing. Many hospital organizations will make collections of their aggregate data available for distribution on various media such as on a CD–ROM or over an Intranet with a mapping front end. ESRI's MapObjects and ArcView Internet Map Servers facilitate publishing of spatial data and deployment of mapping and GIS applications over an Intranet.

Spatial Data Warehousing: Your Map to Success

As data-intensive health care applications grow, the successful hospital organization will need to include the spatial component to its operational data to support the decisions and analyses required to remain competitive. Geospatial data are now being recognized as a powerful visualization and analysis tool in OLAP applications.

For more than 25 years ESRI has been helping people manage and analyze geographic information. ESRI offers a framework for implementing GIS in any organization with a seamless link from personal GIS on the desktop to enterprisewide GIS client/server and data management systems. ESRI GIS solutions are flexible and can be customized to meet the needs of our users.

ESRI is a full-service GIS company, ready to help you begin, grow, and build success with GIS.

Corporate

ESRI
380 New York Street
Redlands, California
92373-8100 USA
Telephone: 909-793-2853
Fax: 909-793-5953

For more information on ESRI software call ESRI at

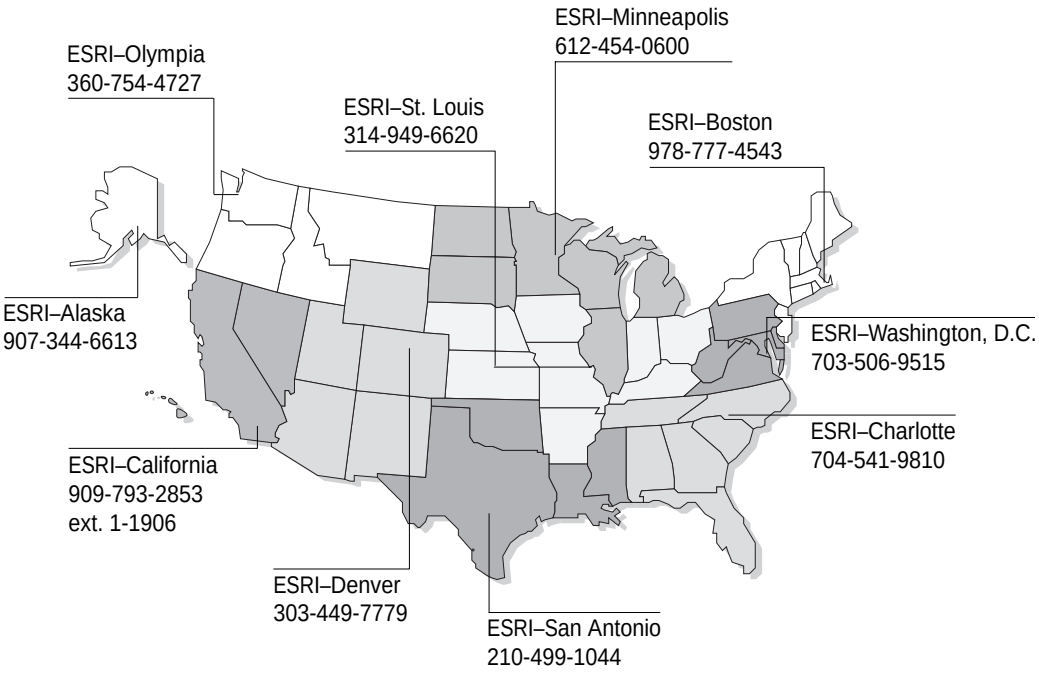
1-800-447-9778

(1-800-GIS-XPRT)

Send E-mail inquiries to
info@esri.com

Visit ESRI's Web page at
www.esri.com

Regional



International

Australia
61-9-242-1005

Canada
416-441-6035

France
33-1-46-23-6060

Germany
49-8166-677-0

Hong Kong
852-2-730-6883

India
91-11-620-3801

Italy
39-6-406-96-1

Poland
48-22-256-482

South Asia
65-735-8755

Spain
34-1-559-4345

Sweden
46-23-84090

Thailand
66-2-678-0707

United Kingdom
44-1-923-210450

Venezuela
58-2-953-0523

Outside the United States,
contact your local ESRI distributor.
For the number of your distributor,
call ESRI at
909-793-2853, ext.1-1235

Printed in USA