Chapter 10 Web-enabled Spatial DB systems

I. Introduction

objective of chapter - explain how spatial DB & internet tech are integrated within the framework of organizational, national, global spatial info infrastructure

2. Definition & characteristics of web-enabled spatial DB systems

2.1 Characteristics of web-enabled spatial DB systems

definition - DB sys designed for access over the internet

purpose - enable external access to enterprise info + support internal business operations such as data sharing among project teams

characteristics & requirements

a. distinct architectures - integrate spatial DB SW tools + server-to-DB middleware + application programming languages + internet markup languages

b. used in conjunction w/ new data collection, processing, dissemination tech

ex. web-cam, weather stations, in-vehicle navigation sys

c. used to serve new business functions

ex. e-commerce, LBS, self-served info kiosks

d. directed toward the widest possible audience

configuration (Fig 10-1)

a. presentation tier - web browser as GUI

b. communication infrastructure tier - LAN, internet

c. business logic tier - contains rules & protocol for interaction between clients & server

web server : connection between the clients & server

web mapping engine : spatial processing in response to client requests

d. data management tier - spatial DB + warehouse + off-line inventory + DBMS

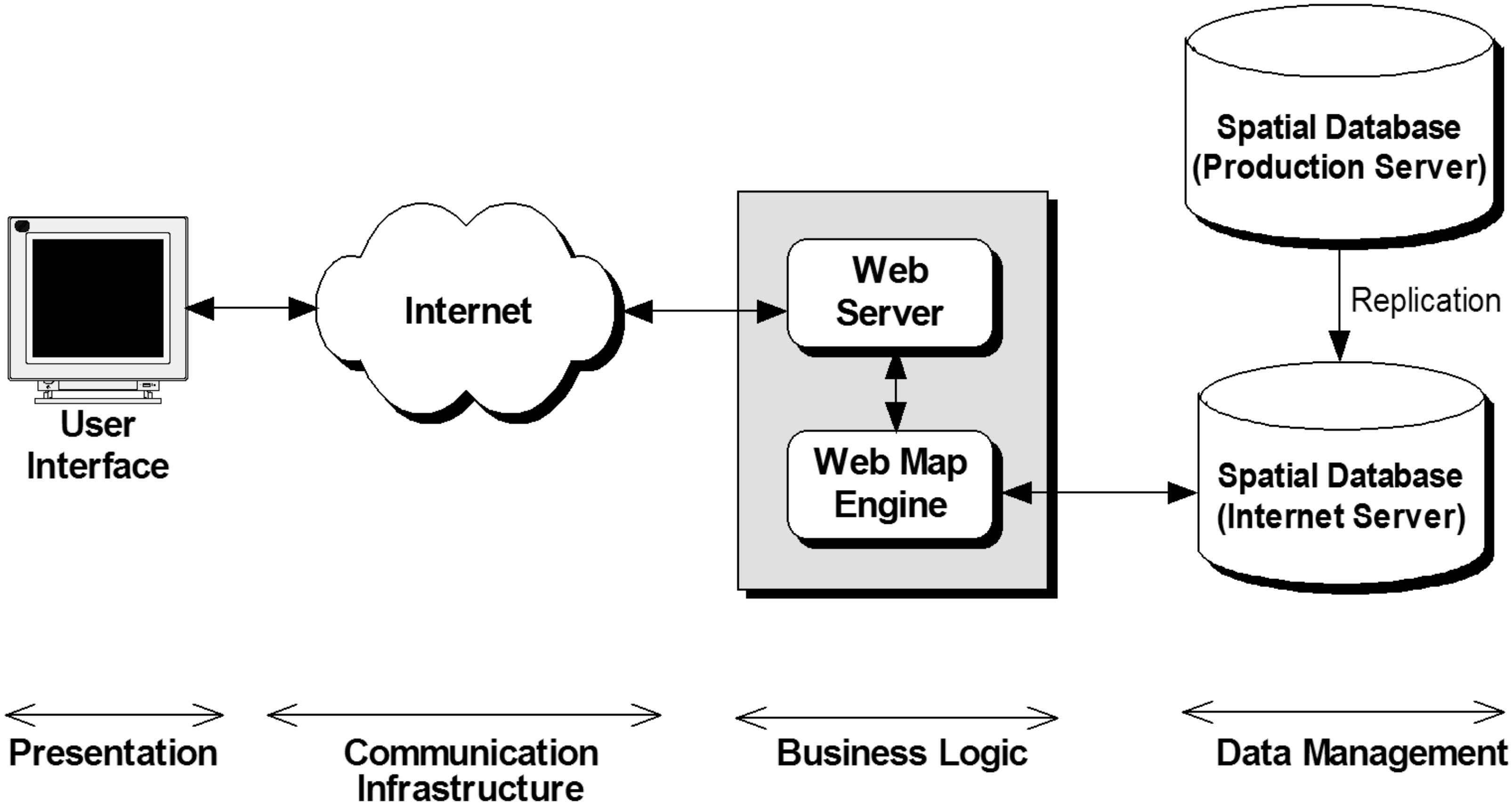

Fig 10-1 Generic architecture of a web-enabled spatial DB sys

2.2 Advantages & disadvantages of web-enabled spatial DB systems

advantages

a. interoperability in a distributed computing environment

by using client/server arch, industry standards, web browser

b. HW & SW independence

DB can be accessed regardless of the computer types

c. rapid development & universal accessibility at manageable cost

as web browser is a standard application installed on all desktop computers

d. lowering the cost of using spatial info technology

ASP(application service providers) offer complete fee-based service packages

e. improved spatial info customer services

provide spatial info service thru world 24 hours, data download electronically, DB is linked to field data loggers

2.2 Advantages & disadvantages of web-enabled spatial DB systems

disadvantage (limitations)

a. inherent limitations of the web

web is designed for text & image based info, limited capabilities for vector data

b. inherent limitations of spatial DBs

DB are heterogeneous - different data models, scales, cartographic + attribute classification schemes

c. security & privacy concerns

data are vulnerable to abuse & can be spied easily

d. copyright control & liability of abuse & misuse

hard to control & retain the copyright of the data, abuse, misuse

3. Technologies & standards

3.1 Info communication over the internet

computers on the internet are referred to as nodes /hosts

connected w/ networks & a set of protocols called TCP/IP (IP address - 32bits binary number)

class - determines domain address & node address (Table 10-1)
  ex. class B : mid-sized network found in a typical university : 16,384 networks for 65,534 nodes/hosts

info is transmitted in blocks called packets, packet contains IP address, size is fixed

IP address is hard to remember -> domain names are used instead
  domain name server breaks down the domain name into IP address

DNS(domain name system) standard
  ex. top level part - com, org, edu, mil, biz / country code kr, ca, jp

ICANN (Internet Corporation for Assigned Names and Numbers) - non profit organization to control them

Table 10-1. Classes of IP addresses

Class of IP Address	***First Octet of an IP Address***	***Domain and Node Parts of an IP Address D = Domain Address; n = Node Address***
A	1 to 126	DDD.nnn.nnn.nnn
	127	Reserved for internal testing on the local machine
B	128 to 191	DDD.DDD.nnn.nnn
C	192 to 223	DDD.DDD.DDD.nnn
D	224 to 239	Reserved for multi-casting
E	240 to 255	Reserved for future use

3.2 Intranet & extranet

intranet - private communication network within an org

- ex. LAN, WAN
- to share an org’s computing & data resources within the org
- in practice, intranet also connects to internet - firewall server is often used to screen

extranet - when part of the intranet is accessible to the users external to the org

- very popular means for business partners to exchange info

3.3 Characteristics of the world wide web (web)

web is developed by CERN (European Lab of Particle Physics) in 1993

an info store on the web is called website/ webserver/ portal

website contains web pages using HTML/ associate markup language

web page is a plain ASCII text file containing HTML commands(tags)

tags define hyperlink - reference other web pages / resources

URL (Uniform Resource Locator) - address of individual page & resource

HTTP (Hypertext Transfer Protocol) - enable access to & exchange of resources on the internet using web

W3C (World Wide Web Consortium) - to coordinate the development of common protocols & tech arch to promote web’s evolution & interoperability

work aggressively to transform the arch of HTTP, HTML, URL to XML based

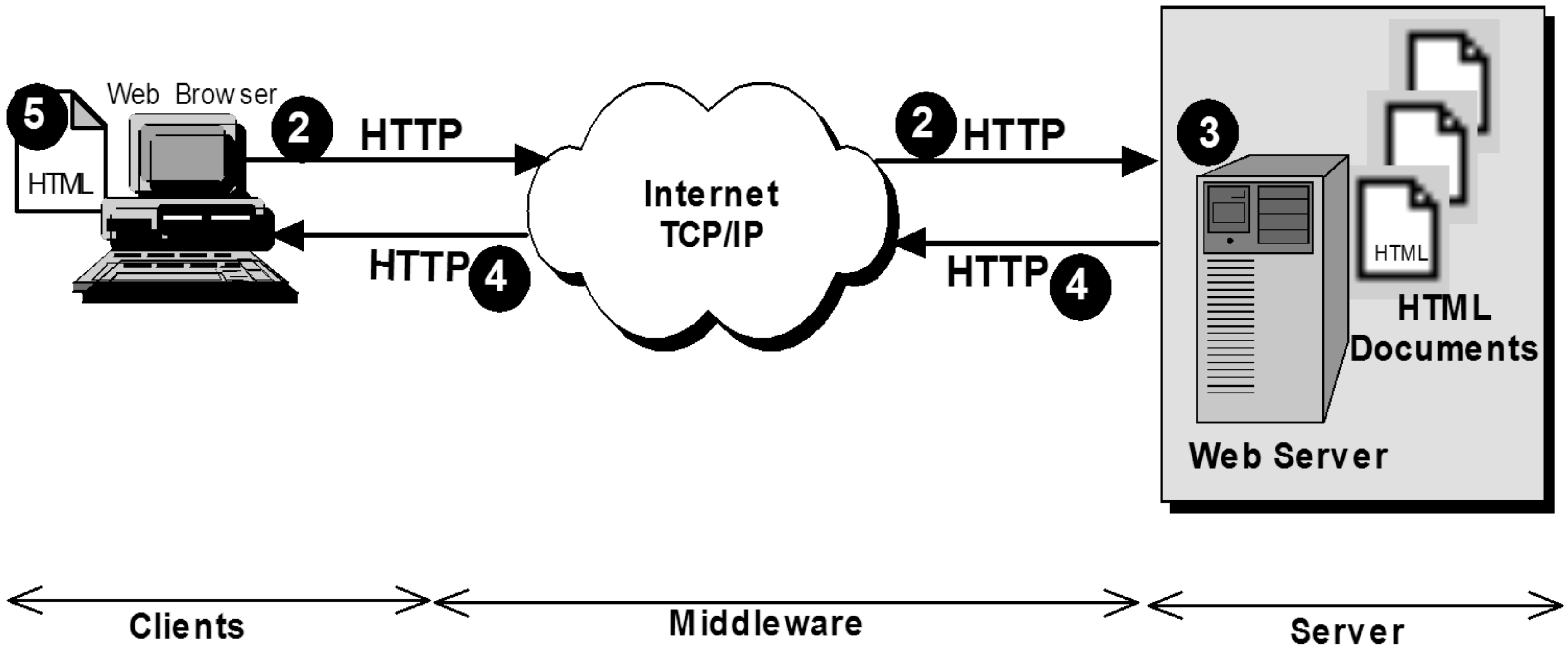

Fig 10-2 Web client/server interaction using HTTP

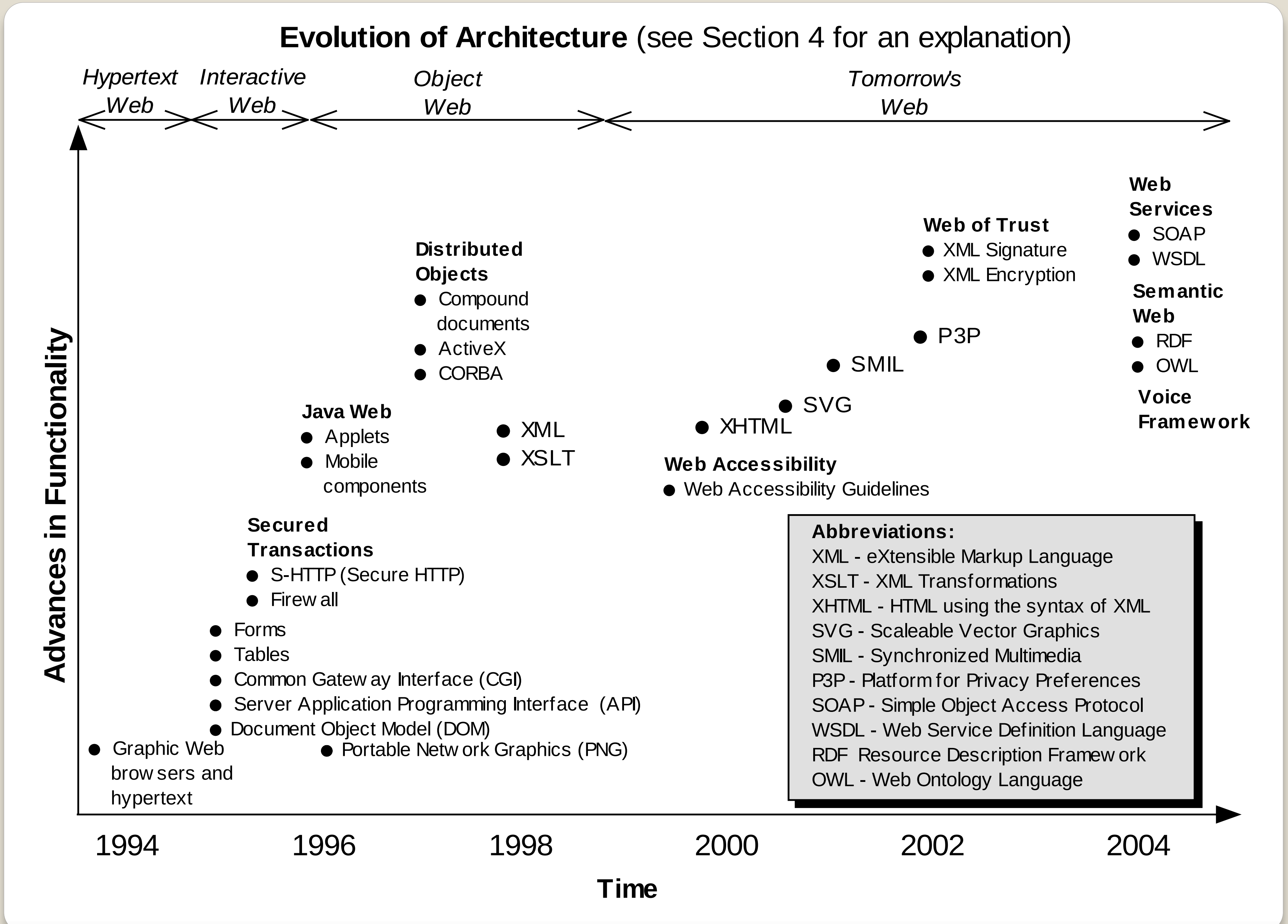

Fig 10-3 The evolution of web technologies

3.4 Extensible Markup Language (XML) & related standards

created by W3C to overcome the limitations of HTML

to curtail the growing complexity of the static, single, all-embracing compendium of HTML tags

permits domain-specific dialects/ grammars, dynamically update web pages thru style sheets

XML family of tech & standards (Fig 10-4)

a. meta-markup language : collection of rules & tags

b. program library tools : to describe & validate the structure of an XML documents, format data, link to DB, query DB, convert documents

c. domain-specific dialects/ grammars

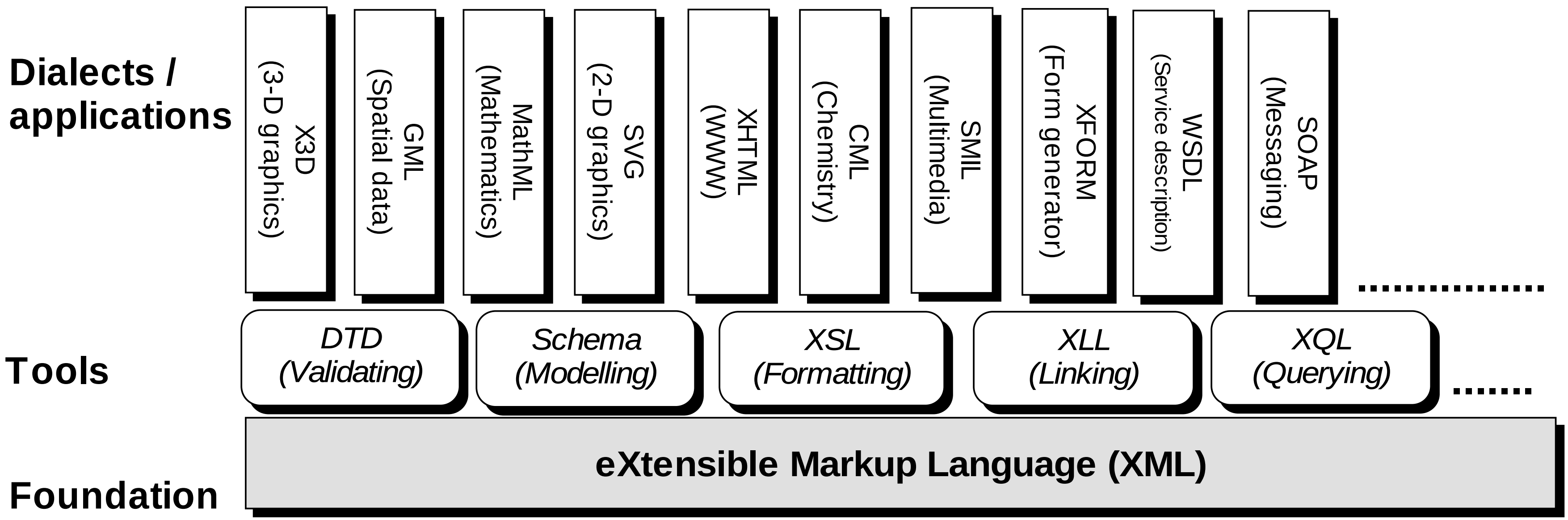

Fig 10-4 The XML family of technologies & standards

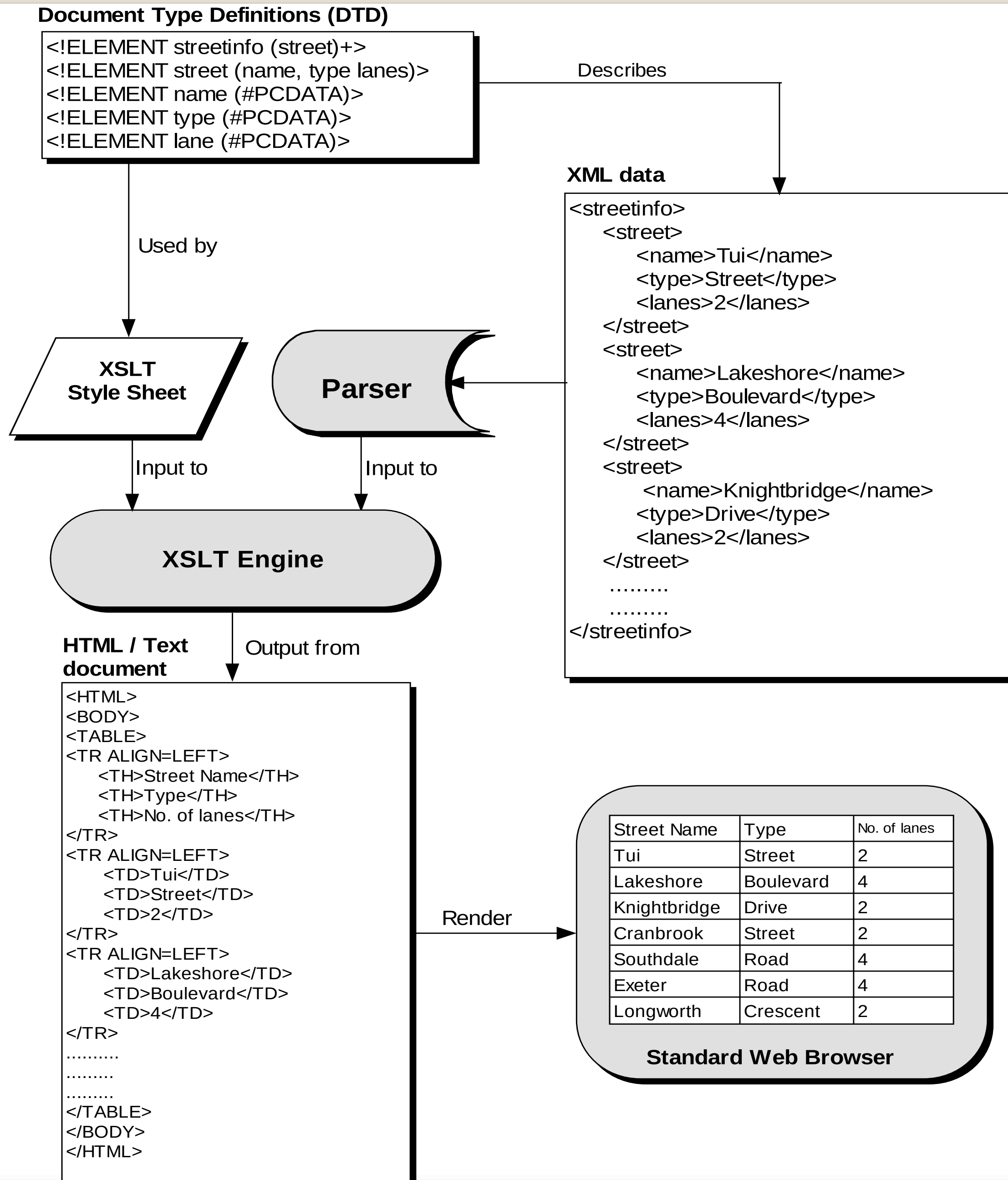

Fig 10-5 Storing, processing & displaying data using XML

for DB application, XML is used in two ways :

a. document-centric model : as a means to create semi-structured documents w/ irregular content

in spatial DB, commonly used to create & deliver meta data

b. data-centric model : as a means to store / interchange format for data in DB (ex. relational DB)

in spatial DB, used to store, transport, share spatial info using the dialect GML

GML - most important milestones for spatial DB sys & spatial DB tech

a dialect of XML, makes full integration of geo info into daily business applications

non-proprietary & based on OGC Abstract Specification -> open source approach

data in GML can be displayed, queried, edited, analyzed at the feature level

SVG - most broadly supported & used output specification

for describing 2-D vector graphic shapes, images, text

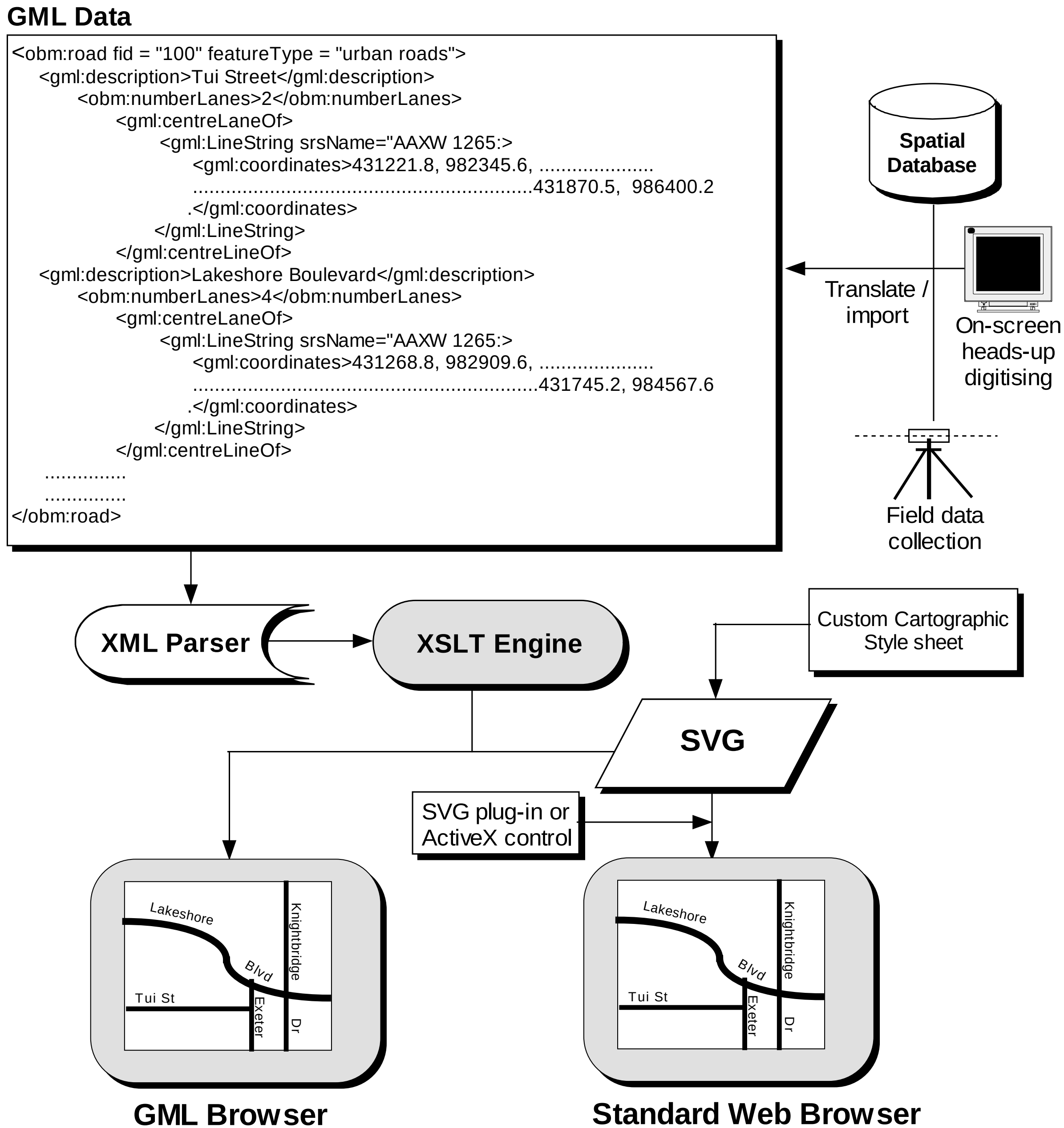

Fig 10-6 Using GML for web-enabled spatial DB applications

3.5 Open Geospatial Consortium (OGC) web mapping standard

OGC specifications cover both data & sys arch - promote interoperability

ex. GML for encoding & transporting spatial data

WMS(Web Mapping Server) Implementation Specification

contains a set of common interfaces for client to query, request, display spatial info from remote spatial DB

generate a map layer in a raster format -> returns a map image file, not original spatial data (Fig 10-7)

WFS(Web Feature Server) Specification

allow to access spatial data in remote DB in object form (points, lines, areas)

provide queries & transactional operations on individual spatial features -> returns results in GML

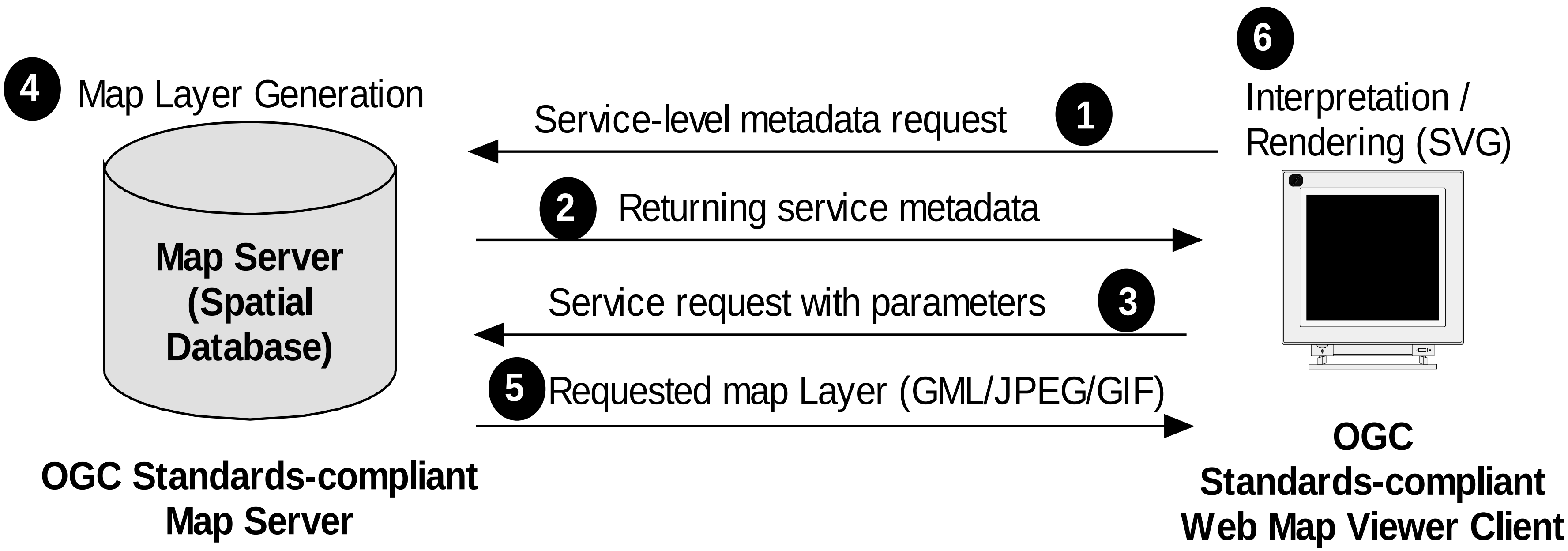

Fig 10-7 Interaction between a map server & a client viewer using the OGC WMS implementation specification

4. Working principles of web-enabled spatial DB sys

4.1 Interactive web architecture

web server is the hurb thru which all accesses to the DB are handled (Fig 10-8)

web server contains a suite of application programs

- ex. web-to-DB middleware : connect a web server to DB, by using CGI(Common Gateway Interface)
  - * CGI : script file to perform a specific function in the web server
- ex. DLL(dynamic-link libraries) : server API, faster than CGI, however server-specific & OS dependent

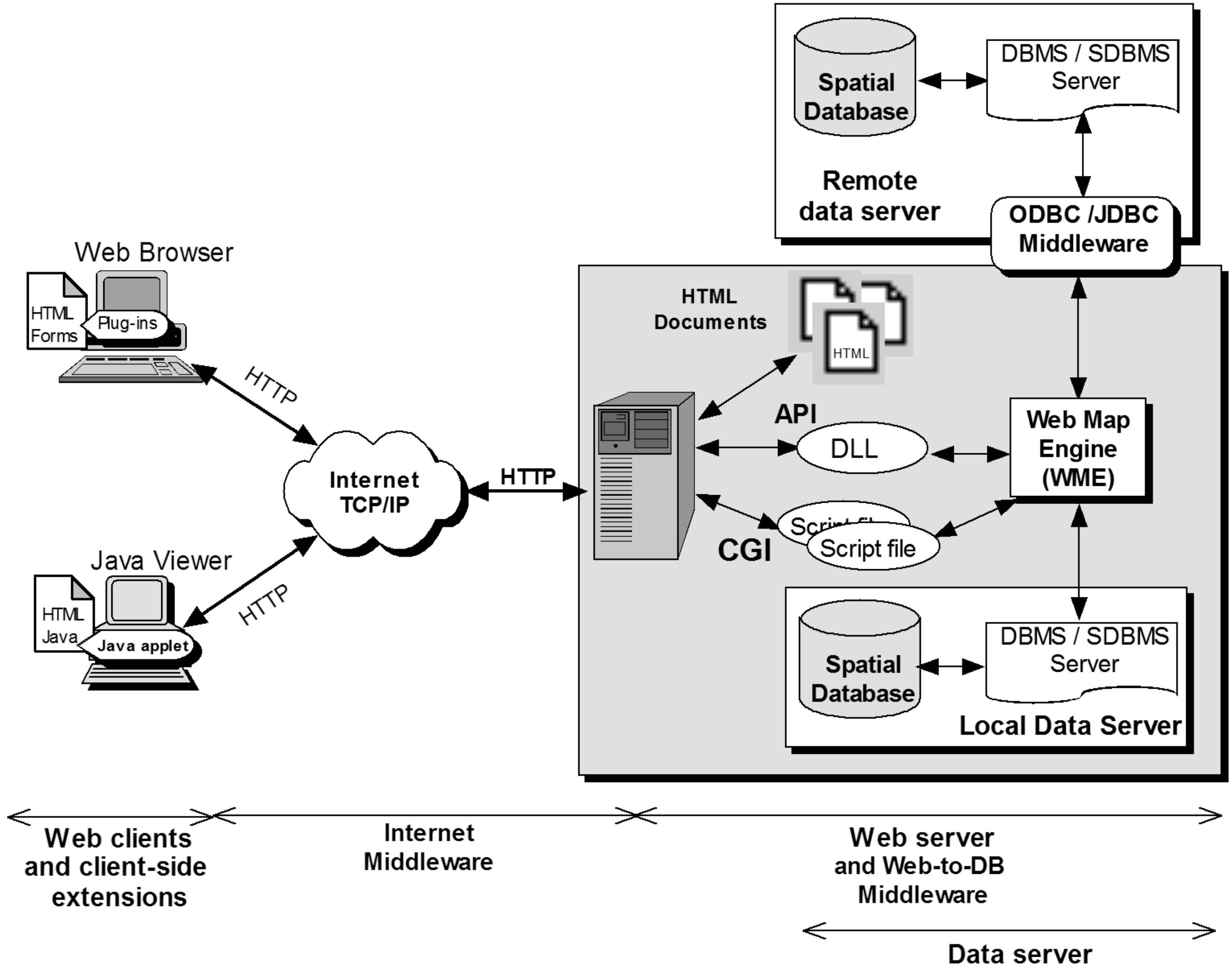

Fig 10-8 The interactive web architecture using CGI & server API

4.2 The object web architecture

three perceived benefits :

- to overcome the limitations of using CGI & server API on the server
- to provide a scalable & robust server-to-server web infrastructure
- to extend Java w/ a distributed object infrastructure

simply put, object web arch = web protocols + distributed object tech

two current competing object web standards :

a. MS DCOM/ActiveX - enables direct communication between clients & servers, bypasses the constraints of HTTP/CGI

b. CORBA/Java of OMG & Javasoft - multi tier client/server application model consisting of Java clients, CORBA business objects, data servers

-> provide platform-independent web-enabled solutions

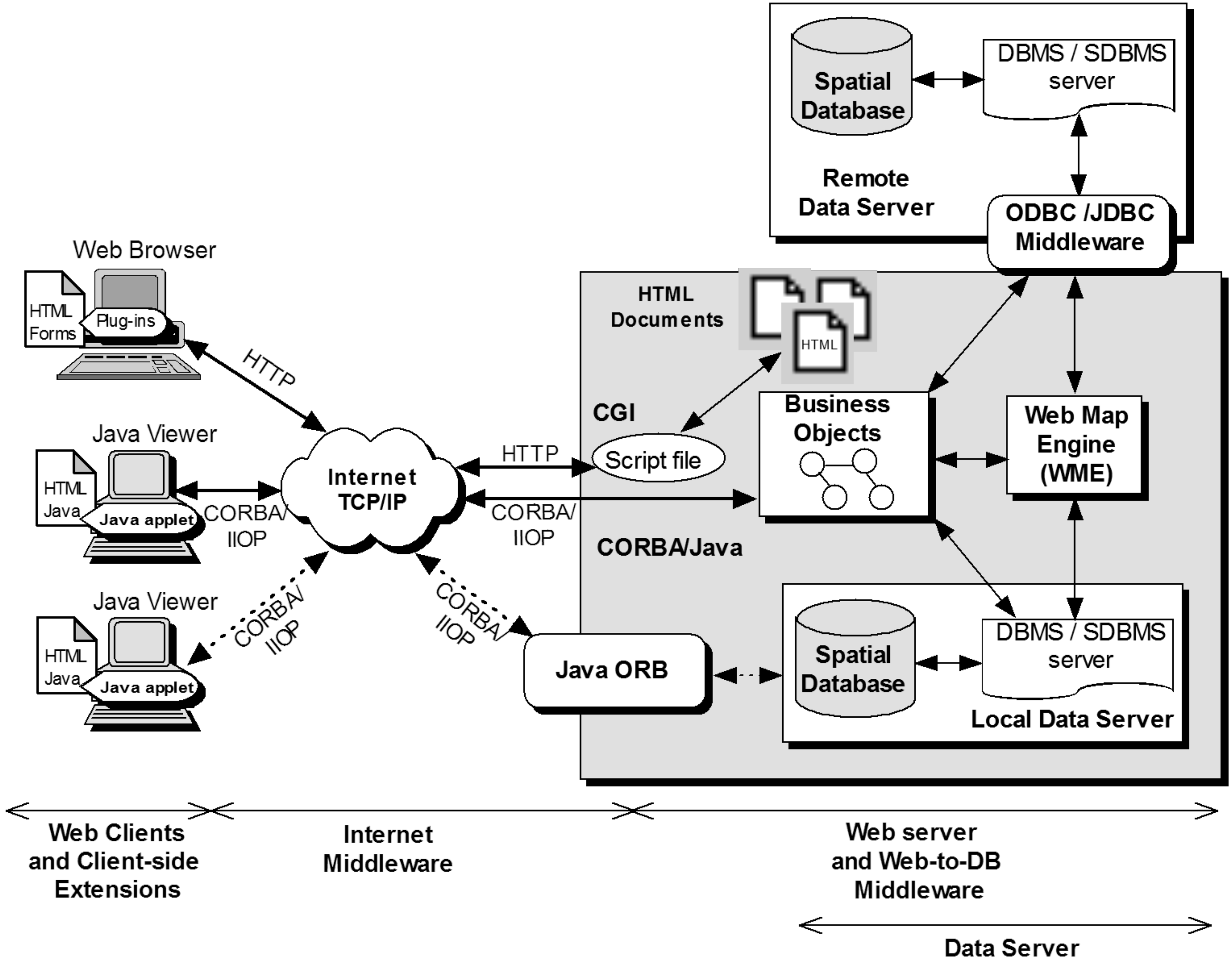

Fig 10-9 The object web architecture based on the CORBA/Java standard

4.3 The next generation web architecture

goal : turn into a medium for large-scale, cross-platform, interactive distributed computing

web services - one of the most broadly embraced concept

- vendor-independent, standards-based means of activating remote cross-platform procedures over internet
- consists of many SW components to build larger, more comprehensive applications (Fig 10-10)
- tripartite relationship - service provider, service broker, service user

service provider - services are built on an arch made up of a stack of SW layers

- at the foundation - standard such as HTTP, TCP/IP, HTML
- next layer - SW tools + interfaces
  - ex. XML - basic foundation, provides a language for defining data & method of processing them
    - SOAP(Simple Object Access Protocol) - a collection of XML based rules defining the format of communication between a web service & clients
    - WSDL(Web Service Description Language) - another set of rules defining a web service interface, data & message types, interaction patterns, protocol mapping
    - OWS(OpenGIS Web Services) - OGC family of standards
- top layer - service layer, consisting of spatial data processing programs

service broker - catalog service/ clearinghouse

- UDDI(Universal Description, Discovery and Integration) - web service registry & discovery mechanism used for storing & categorizing business info & for retrieving pointers to web service interfaces

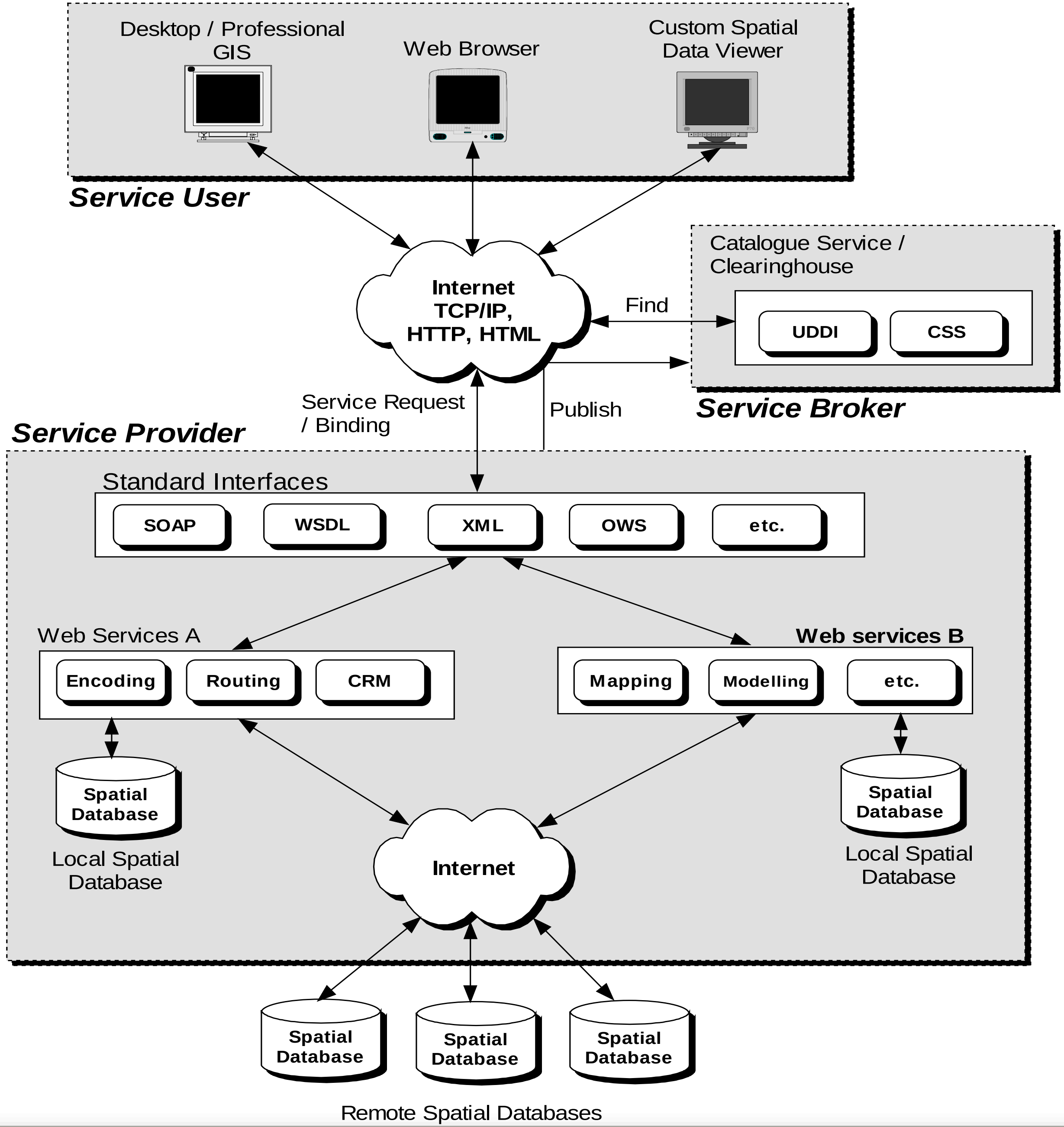

Fig 10-10 Delivering DB applications using web services model

semantic web - by Tim Berners-Lee at the W3C(2001)

- to address the limitations of current web
- by augmenting HTML/XML documents w/ an additional layer of metadata & logic rules
- possible by a set of standards - RDF(Resource Definition Framework), OWL(Web Ontology Language)

semantic geospatial web - by Egenhofer(200)

- represent semantics using multiple spatial & terminological ontologies

5. Implementing a web-enabled spatial DB sys

5.1 Design considerations

- there are a number of factors that govern the direction & success of the web-enabled spatial DB sys
  - grouped into 8 categories (Table 10-2)
  - must be considered in tandem

Table 10-2. Design and implementation requirements

Primary Factors	***Secondary Factors***
Purpose of implementing the spatial database system	○ Mission and business goals of the organisation. ○ Providing the general public information versus supporting transactional processing in business operations. ○ Delivering maps versus providing attribute spatial data. ○ Providing maps versus interactive querying. ○ Browsing versus downloading. ○ Data viewing versus data analysis. ○ Internet versus intranet versus extranet.
Audience or users of the system	○ Internal (same or different locations) versus external users. ○ Level of computer literacy and skills. ○ Exposure to spatial database technology. ○ Accessibility to the Internet.
Technology supporting the system	○ Server requirements (UNIX-based versus high-end PCs). ○ Client computer requirements (thick versus thin clients). ○ Bandwidths of local and wide area networks. ○ Plug-n versus Java applets versus Active X controls. ○ Web browser compatibility. ○ Graphical user interface. ○ Integrations with other technologies (e.g. GPS).
Architecture of the system	○ Operating systems. ○ Security and firewall requirements. ○ System Architectures (see Sections 4.1 to 4.4). ○ DBMS software, Web server and Web Mapping Engine. ○ Off-the-shelf applications versus custom applications.
Spatial data to be served by the system	○ Raster versus vector data. ○ Possible applications. ○ Standards and quality. ○ Sources and volume of framework / application data. ○ Intellectual property and copyright. ○ Updating frequency.
Application requirements and constraints	○ Business logic and processes. ○ Work flow. ○ On-line Analytical versus On-line Transaction Processing.
Cost of capital investment in the system	○ Cost of hardware / software. ○ Cost for systems operation and maintenance. ○ Cost to clients. ○ Free distribution of data versus commercial selling of data.
Legislation and regulations for collection, access to and application of spatial data	○ Domain-specific federal and provincial/state legislation. ○ Domain-specific federal and provincial/state regulations. ○ International, national and provincial/state data standards. ○ Multilevel hardware/software standards. ○ Communication and telecommunication infrastructure.

5.2 Approaches to implementation

- info centric vs application-centric approach
  - info centric : rapid access to info is the primary concern, ex. MapQuest, GoogleMaps, MS Virtual Earth
  - app-centric : provide domain-specific data for pro users, ex. KLIS
- org vs infrastructural approach
  - org approach : serve business needs of an org, for well-defined user community
  - infra approach : provide basic topo & socio-economic info to the general public
- in-house development vs outsourcing
  - in-house : from conception thru design to implementation by the org
  - outsource : take SW options to be purchased, customize
- client-oriented vs server-oriented approach (Table 10-3)

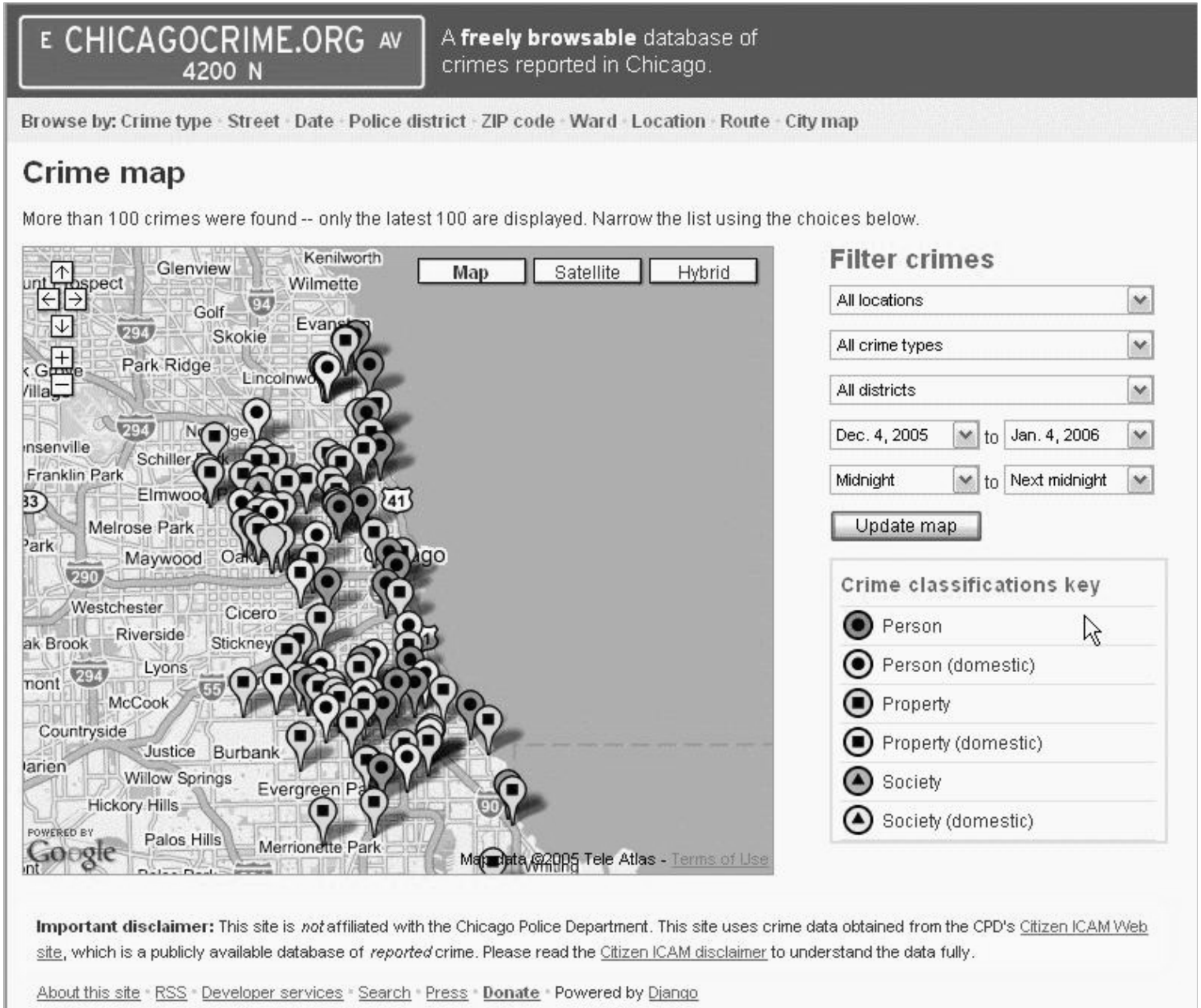

Fig 10-11 Example of a web-based spatial DB application hosted by an application service provider

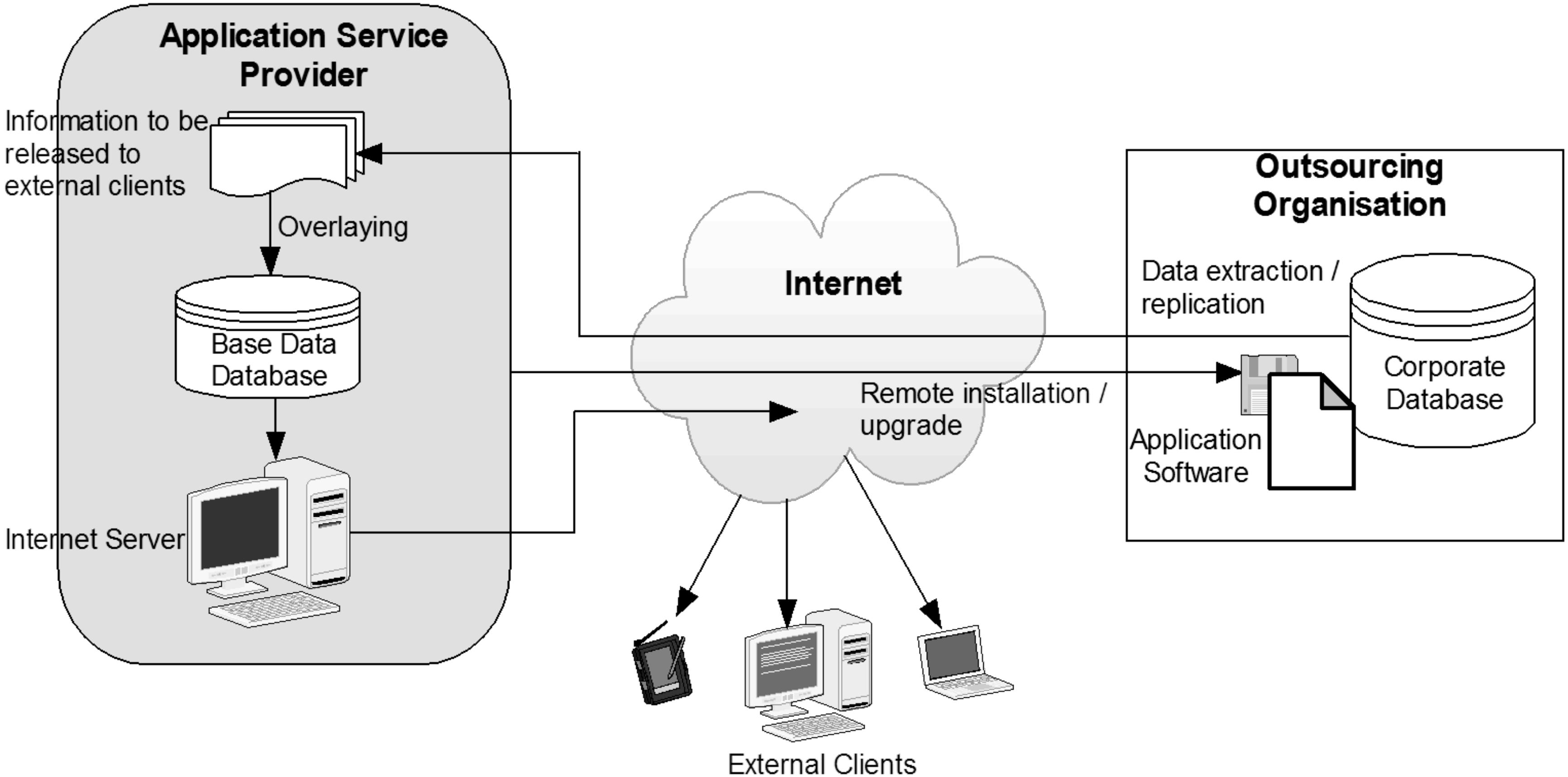

Fig 10-12 The portal model of an application service provider (ASP)

Table 10-3. Comparison between server-side, client-side and hybrid strategies to web-based spatial database implementation

Strategies	*Advantages*	*Disadvantages*
Server-side	o Allows users to access large and complex data sets that are otherwise difficult to transmit across the Internet. o Uses a high performance server capable of handling sophisticated computations in complex applications. o Centralised control over data and processing to ensure applications are correctly deployed. o Possible lower overall implementation cost since there is no need to install costly hardware and software at every user site.	o Slow processing since each data access and processing step must be returned to the server for execution. o Performance is subject to available bandwidth and network traffic on the Internet between the server and the client. o Inability of applications to utilise fully the processing power of modern desktop and laptop computers.
Client-side	o Inability of applications to take full advantage of the processing power of the client computers. o Greater user control of data processing and analysis in applications, thus allowing interactive computing where intermediate results must be interpreted before the next processing step is taken. o Better data processing due to lesser needs for client-server communication and data transmission.	o The requirement to transmit large amounts of data from the server to the client(s) may cause delays. o Large and complex data sets may be hard to store and process on less powerful client computers. o Sophisticated analytical and modelling applications may run more slowly on less powerful client computers. o Sophisticated analytical and modelling applications often require expensive training in order to use them properly.
Hybrid	o Can make full use of the relative advantages of server-side strategies while minimising their relative disadvantages. o Can make full use of the relative advantages of server-side strategies while minimising their relative disadvantages.	o Difficult to determine a balanced division of work between server and client computers because of variability in the nature of applications, the processing power of the computers used, the technical background and training of the users, and the bandwidth that is available to access the server.