



Distributing proprietary geographic data on the World Wide Web — UCLA GIS Database and Map Server

Yafang Su*, Joan Slottow, Avi Mozes

Office of Academic Computing, University of California, Los Angeles, PO Box 951557, CA 90095-1557, USA

Received 26 April 1999; received in revised form 30 September 1999; accepted 30 September 1999

Abstract

The University of California, Los Angeles (UCLA) GIS Database and Map Server project is a cooperative effort of GIS-interested-parties on the UCLA campus, to deliver centrally-stored geographic data via the World Wide Web (WWW). Both geographic data and geo-referenced remotely sensed image data are served from this site. The main objective of this project is to serve gigabytes of data using limited online disk space, used to store the most-recently-requested data units, and a large tape robot for the offline storage of all the data in the central database. ArcView Internet Map Server (IMS) is employed as the interactive GIS mapping server and visual interface for downloading the data from the central database. The server design, database design and web client interface were developed to incorporate a visual interface, to distribute proprietary geo-data and to provide GIS capabilities. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Internet GIS; Map server; Database; Digital library; World Wide Web

1. Introduction

Since the first web page with an interactive map was put up by Xerox Palo Alto Research Center (PARC) as an experiment in interactive information retrieval in 1993 (Harder, 1998), sites have increasingly appeared on the web serving maps, geographic data, or both. Plewe (1997) discussed the subject of GIS online and listed many sites that were distributing data. Among the data downloading sites [1–15], some sites provide public accessible geographic data. For example, the USGS provides geo-data and remotely sensed data online [14]; California State University Northridge pro-

vides some publicly accessible geographic data [4] and the Alexandria Digital Library site [1] has many links to significant collections of geo-spatially referenced information. Many sites from data vendors provide commercially available geo-data for users to purchase such as Microsoft's TerraServer [15], and the Teale Data Center's geographic data [10]. Many interactive GIS mapping sites [16–30] are publicly accessible and some of them are very useful and close to people's daily life. MapQuest [24] provides a good site for online maps and directions within the United States (US). National Real Estate Finder (Reality.Com) [27] allows users to search over one million new and resale homes in US. All the vendors of Internet GIS products have demo sites which also allow users to do interactive mapping, such as ESRI [20] and MapInfo.

GIS is becoming a popular tool in many fields. Although Microsoft's MapPoint 2000 is not a high-

* Corresponding author. Tel.: +1-310-206-3702; fax: +1-310-206-7025.

E-mail address: yafang@ucla.edu (Y. Su).

end GIS, it will ship with its Office 2000 suite (Bal-tazar, 1999). That will make GIS more accessible to the public.

Geographic data is one of the most important parts of a GIS system. Almost all universities have purchased some data from software vendors or data vendors. Although the US and some other western countries provide government-owned data free to the public, there are many other countries and organizations that consider their data to be proprietary. To be cost effective, organizations such as UCLA usually buy these data for organization-wide use. Therefore, methods to maintain a license agreement and share proprietary data within an organization becomes an important issue, especially the distribution of data within the community via the Internet.

The UCLA GIS Database and Map Server web site (URL: <http://gisdb.cluster.ucla.edu>) was developed by the UCLA Office of Academic Computing and is currently supported by six interested partners. It is an effort by those parties at UCLA interested in GIS to serve gigabytes of data using a hierarchical data storage management system (Tape Robot) to store centrally-distributed data. A committee composed of seven faculty and staff members reviewed the proposal and prototype web site created for this project.

The UCLA GIS Database and Map Server is both a data distributing and interactive GIS mapping site. We developed a visual interface for downloading the data, which also allows users to view a subset of the data and interact with that subset on the Internet using a limited set of GIS tools provided by the server.

On our web site, we also direct users to many other public web sites by maintaining links for interactive GIS mapping and geographic data downloading. Internet references [1–30] are some of the selected links.

This site promotes GIS on the UCLA campus and encourages the UCLA community to use geographic data and GIS. We have also added information about the UCLA-ESRI site license agreement to the site. We include links to other sites from which users can download freeware such as the ArcExplorer Data Browser from ESRI, Redlands, California and data format conversion shareware such as Geographic Explorer from Blue Marble Geographics, Cardinar, Maine. After users download the data from the UCLA GIS database to their local machine, they can use ARC/INFO or ArcView software obtained via the site license, the freeware or other software to convert, view or explore the data.

We discuss the issues related to the distribution of geographic data on the World Wide Web and share the experience of developing our site in this paper.

2. Objectives

The main objectives of the web site are to:

1. Provide proprietary data to the UCLA community and a channel for the members of the UCLA community to share their data within a department or departmental group.
2. Store all the data on a large Tape Robot and retrieve it for downloading on demand.
3. Develop a visual interface for downloading the data.
4. Provide a publicly available interactive GIS mapping site.

2.1. Proprietary and limited-use data

This site protects the data from illegal use by only providing specific data items to those who are licensed or otherwise allowed permission.

The site allows members of the UCLA to obtain new UserIDs by filling in a form online at the site. An immediate check validates the user as a member of the UCLA community, and a UserID is added for him/her, which is assigned to a particular user group. Access to data belonging to a restricted group requires that the user contacts the user-group administrator.

UCLA departments that provide data to be added to the site can control access to those datasets by assigning them to restricted groups. Only the administrators in the departments can add group access for departmentally-owned data to individual UserIDs.

2.2. Visual interface for downloading the data

The interface allows:

1. Users to select the geographic data of interest by clicking either on lists of available data (which are generated on the fly from information in the Database Catalog) or on maps. This allows users to select geographic data to download by viewing maps in conjunction with lists.
2. Users to view a pre-selected subset of the data and interact with that subset on the Internet using a limited set of GIS tools powered by ArcView Internet Map Server. For this purpose, we make available those maps needed to provide the user interface and additional sample maps that allow users to work with a subset of data from each data category.
3. The public to access the site, and interact with the sample maps in order to learn some basic GIS capabilities and to produce some maps from the pre-selected data available on the site. Only when the user requests to download the data, is he/she asked

for a UserID. This protects the proprietary nature of the data and limits the downloading to the UCLA community.

3. Web server design and database

3.1. Web server design

The UCLA GIS Database and Map Server web site is a publicly accessible site. We use the Netscape Fast-Track server as our web server. For most sites, one server is adequate. But since we are serving two audiences, (the general public and the UCLA community) we have two servers running on this site (see Fig. 1). The public server allows access without authentication. The other requires UserID verification. Users are passed automatically from the public server to the internal server when they request that data be downloaded. The twin server architecture is to protect the proprietary nature of the data.

3.2. Database catalog

The database catalog stores all the information about the data in our central database. It drives

the user interface and determines the way of searching the database. The fields of the database catalog are shown in Table 1.

The five location fields (Country, State/Province, Location1, Location2 and Location3) and the three location type fields in the database catalog are for searching the database by geographic location. An index database catalog, containing only one record with each unique combination of those entries, significantly accelerates the query process.

The two data unit fields (Data Unit 1 and Data Unit 2) and the two data unit type fields are used for raster data or image data because those data are usually subdivided into grids or satellite scenes. Before a user finally downloads the data, he/she is presented with a brief metadata display made up of the contents of the data theme, source, format, scale/resolution, projection, filesize and datadesc fields. Reading this information helps the user understand the data content and decide whether or not to download the data. The filename and filesize fields are for online data management.

The restriction field is for the program to decide if the user has the permission to download the data. Most of the data are restricted to the UCLA community.

The fields listed in Table 1 are designed for querying

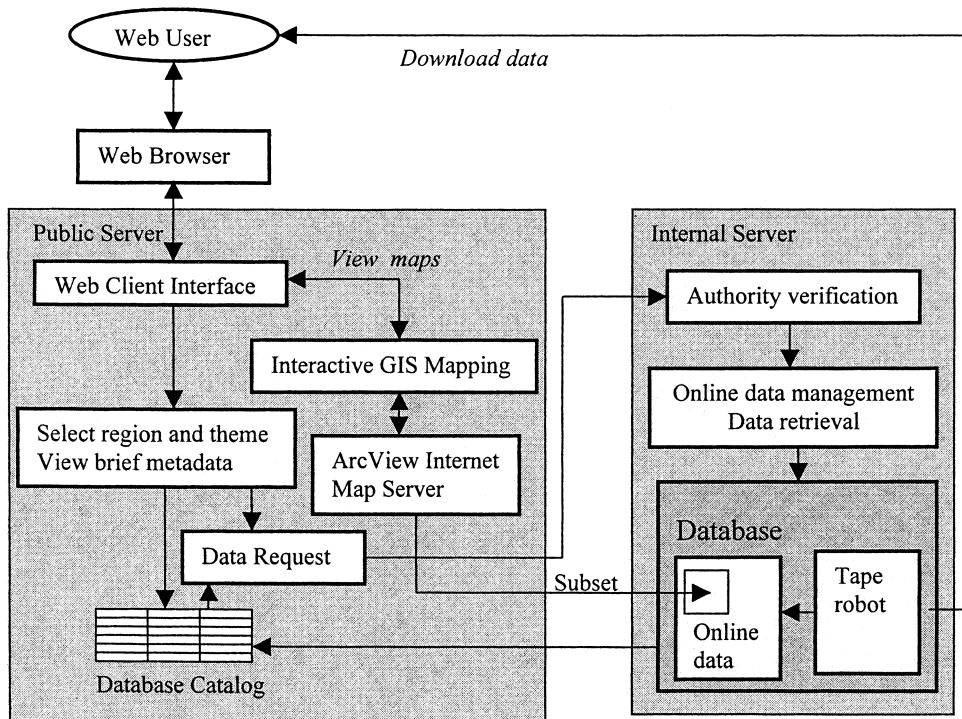


Fig. 1. Diagram of web servers and their functions.

from the web interface. Most data are first searched by geographic area because users usually know the location of the area of interest. GIS data are available at different scales and levels of detail, and users can select maps at an appropriate resolution.

3.3. The central database

The central database behind the UCLA GIS Database and Map Server contains geographic data and image data. Currently, all the geographic data are stored either in Arc/Info coverage, Grid format, or ArcView Shapefile format because the original data came in those formats. Most GIS software can read or convert these formats. The image data are stored in some standard formats such as Geotiff. The data are broken up into manageable units for downloading. Each data unit is served to users as a complete package in a single tar or compressed tar file. Each package contains the data itself, detailed metadata, a README file and other supporting files such as an Arc Macro Language (AML) program to extract the data. Metadata can be considered as the biography of the geographic data (Zurier, 1996). Since the Federal Geographic Data Committee (FGDC) first released the metadata standard in 1994, metadata has become an

integral part of geographic data, providing information about the accuracy or quality of the data.

Currently the database includes Teale geographic data and Landsat Thematic Mapper (TM) images, ESRI's data, some of the Tiger street and address 1994 data, aerial photo data for the city of Los Angeles, Wessex 1997 Tiger data, and others. These data are requested by and licensed to UCLA.

3.4. Data retrieval from the tape robot

A Tape Robot is an automated digital tape library. Each library can store thousands of tapes and each tape can store 10 GB of data in uncompressed or 30 GB in compressed form. Many organizations use it as a daily backup for their system and off-line data storage. A Tape Robot can have several tape drives. A mechanical "arm" selects the tapes and places them in the drives. Each drive can transfer data at a rate of 9MB/s to 14 MB/s. Our Tape Robot is managed by IBM ADSTAR Distributed Storage Manager. Currently, although the disk space is getting progressively less expensive, to have the many terabytes of online disk space that are required for a large GIS server is still a significant cost.

Much of the programming effort required to bring

Table 1
Database catalog fields

Field name	Description	Example
Country	Country or region name (put. for World)	USA
State or Province	Country sub-unit name such as state, province, etc.	California
Location 1	Further sub-unit name such as county, parish, etc.	Los Angeles
Location 1 Type	Further sub-unit type such as county	County
Location 2	Further sub-unit name such as city, mainland, offshore	Santa Monica
Location 2 Type	Further sub-unit type for Location2	City
Location 3	Further sub-unit name: expandable field for future	
Location 3 Type	Further sub-unit type for Location3	
Data type	Data types such as geographic, TM, DEM	Geographic
Data Unit 1	Further data sub-unit name such as scene no.	5230218050
Data Unit 1 Type	Further data sub-unit type for Data unit1	Scene
Data Unit 2	Further data sub-unit name: expandable field for future	
Data Unit 2 Type	Further data sub-unit type: expandable field for future	
Data Theme	Name of data theme	CensusTracts
Map Covers	Area covered by map: 0 = world, 1 = country, 2 = state, 3 = location1, 4 = location2, 5 = location3	4
DataDesc	Brief description of data theme	1990 Census tracts boundary
Source	Where data comes from	TEALE
Format	Format of data	ARCINFO Cov.
Scale/Resolution	Scale or resolution of data	1:100000
Projection	How data are projected	Geographic DD
Restriction	Group restriction — name of group	UCLA
Filename	File name of data in Tape Robot	Teale.usa_ca_..censustracts.tar
DiskSpace	Size of data set in Megabytes	19.23

this web site to fruition has involved the management of disk space and tracking of requested files.

3.5. Online data management

About 16 GB of disk space are currently available for online data and more will be added in the future. When a user requests a file, if the file is online, the file is immediately downloaded to him/her. If however, the file is not online, the request is recorded and the file is retrieved from the Tape Robot. The user can return in about 10 min to download the file. A user can check the status of requested files using a utility program.

Requested files are tracked in a table that also lists those files currently online. A last-time-accessed timestamp is maintained for each online file, and the UserID and the time of request are kept per file for each requestor who has not yet returned for it. A program which runs periodically looks in this file for the names of those requested files which are not yet online and it retrieves them from the Tape Robot. It also automatically sends an e-mail message to the user informing him/her of the file availability. In addition, this program manages the disk space, using the timestamps to determine which files should be deleted if the online disk is full.

3.6. Administrative tools and utilities

Administrative tools and utilities have been written for use by departmental administrators to manage membership in their groups. The administrative utilities include: create a new user, add an existing user to a group, show your users, and show your groups.

4. Choosing the right “Web Widget”

4.1. Map server

The map servers available from major GIS vendors include: Autodesk MapGuide, ESRI ArcView Internet Map Server (IMS), ESRI MapObjects Internet Map Server, Intergraph GeoMedia Web Map, and MapInfo ProServer and the current release of MapXtreme. Limp (1997) and Culpepper (1998) reviewed these products. The packages differ in: (1) the use of image graphics vs graphic objects, (2) the use of map pre-publishing vs direct data access, (3) their client-level requirements, (4) the amount of geographic analysis accessible from the web, and (5) data and platform restrictions.

The web’s strength is its platform-independent nature and interoperability, which allows users with browsers on various platforms (Macintosh, UNIX and Windows) all to see the same thing. However, this is

not the situation with Internet map server products. Some have browser requirements that limit the platforms supported, and even greater restrictions exist on the server side. At this time, all of the authoring/server parts of the map server packages listed previously require a Windows-based system, except that ArcView IMS has both Unix and Windows versions. The recently released MapXtreme Java edition runs on all platforms which support Java.

The GIS functionality of the listed map servers is similar, however the data formats supported by them are different. Each product favors its own proprietary data format, requiring that maps to be made available online must be in the appropriate format.

To choose a suitable map server, a user must match a map server and an application in the following areas: data format used for geographic data, web server platform, GIS functionality, and pre-publishing or direct data access. The UCLA site has ArcView IMS and MapObjects IMS obtained via an ESRI site license, and MapGuide Server obtained from Autodesk. The MapGuide Server runs only on PCs, and it only supported PC browsers at the time this project was under development, since it required a Windows-only plug-in on the client side. Now the plug-in also supports Unix and Macintosh platforms. Newer map servers will make web GIS applications more flexible and capable than ever. For example, MapXtreme Java edition released by MapInfo, provides the ability to create intelligent web pages that allow customers to find information dynamically (Baltazar, 1999). ArcIMS 3.0, developed by ESRI, features an out-of-the-box solution for creating, designing, and managing web sites that incorporates mapping and GIS. With ArcIMS, local data can be integrated with Internet data in a simple browser interface giving users GIS capabilities in an easy-to-use framework (ESRI, 1999).

The UCLA web server platform is Unix and most of the geographic data in our current database are in ARC/INFO coverage, ARC/INFO grid or ArcView shape formats. We prefer the data pre-publishing rather than direct access method of working, because of data license restrictions. ArcView IMS is the out-of-the-box map server available for UNIX. It also meets our other requirements. Therefore, we have chosen the Unix version of the ArcView IMS as the map server for this web site. Generally, the developers should select an IMS product which can best incorporate their organization’s hardware and software environment and best perform their tasks.

4.2. Client interface design

HTML, the C programming language and ArcView IMS were used to develop the client interface. HTML is used to make multi-frame interfaces. ArcView IMS

is used as the map server. C programs are used to query, retrieve and manage the data.

For users to locate the data in our database, we provide two versions of the interface:

1. The Interactive GIS Mapping version allows for some interactive GIS mapping powered by ArcView IMS. Here users can click on maps or lists to go to the area of interest and download the associated data.
2. The Sample Static Maps version is for those with non-Java-enabled browsers to see sample maps and download the associated data. Those who do not want to do interactive mapping can also go to this version to download the data quickly.

The UCLA GIS Database and Map Server front page (URL: <http://gisdb.cluster.ucla.edu>) acts as an entry point to the above two versions of the interface and the page from which users can download previously requested data.

This home page also provides services for requesting a UserID, changing a password and viewing the information about the data in the database. After clicking

on one of the Starting Maps in each version's home-page to start the map search, the user is presented with the five frames that make up the main client interface (Fig. 2). All the frames are scrollable and resizable.

1. Frame 1: map starting points: this is a clickable list of the maps available in the visual interface arranged by category. This frame is constant throughout the site and is always displayed on the left side of the web page. It displays the names of two kinds of maps: the Starting Maps and the Grid Maps. The Starting Maps serve as the basis of the user interface. Each map contains geographic boundaries as well as several sample layers of data. The Grid Maps are for DEM and image data, which is divided into grids. This frame is always displayed unchanged.
2. Frame 2: the title of the map.
3. Frame 3: the main mapping frame including tools, maps, the status bar etc. In the Interactive GIS Mapping version of the interface, an interactive map with mapping tools is displayed here (see Fig. 2) for geographic data and a clickable map (see

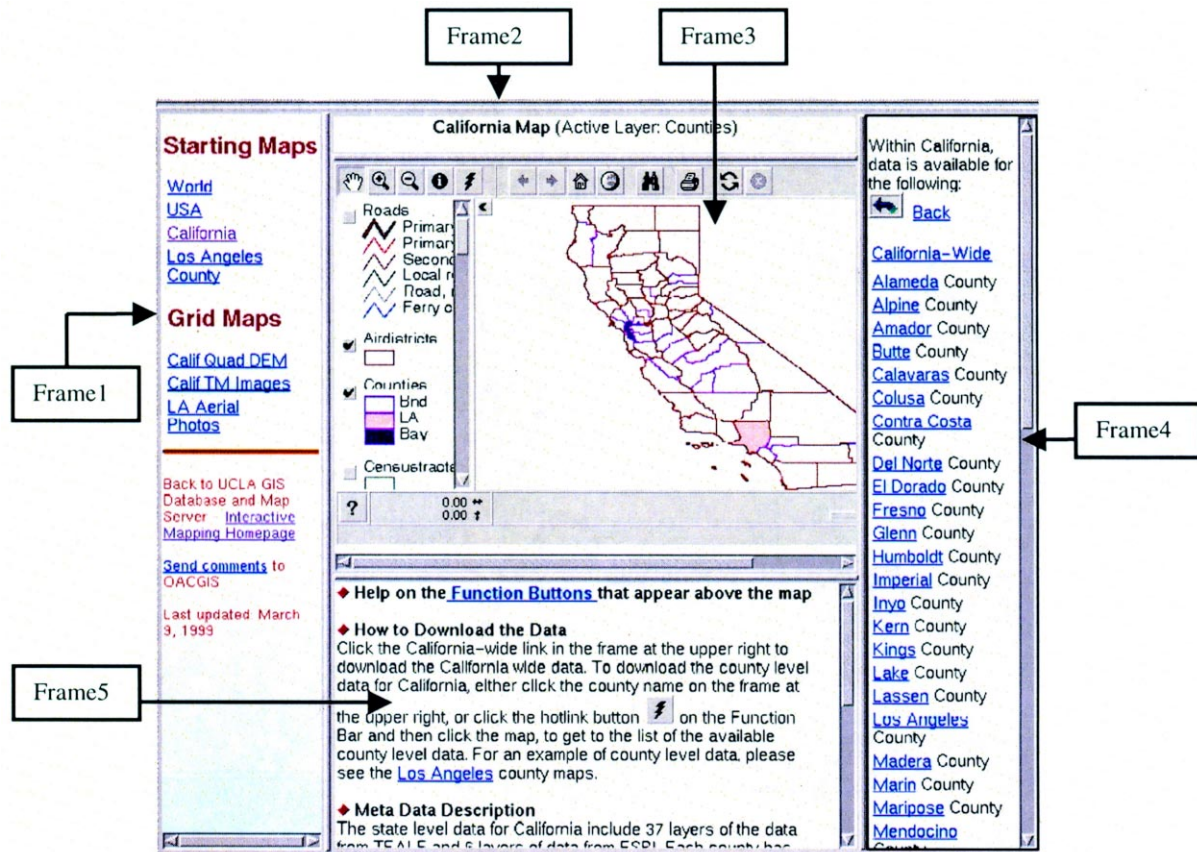



Fig. 2. Example of interactive mapping and data downloading page.

Fig. 3) is displayed for grid data or images. In the Sample Static Maps version of the interface, a static map replaces the interactive map for geographic data.

4. Frame 4: the data downloading interface for the authorized community. This frame lists the data available for the area displayed in Frame 3. It allows all users to browse the brief metadata, and those users who have UserID to download the data. In the example shown in Fig. 2, if the user clicks on California-wide or one of the county names, the list in Frame 3 will change to the list of the types of data such as geographic or Landsat TM available for that geographic location. A further click will go to the list of the available themes for this data type. In the Interactive GIS Mapping version, the user can download the county level data for California, either by clicking the county name in this frame, or by clicking the hotlink button  on the Function Bar and then clicking the map on the county of interest. Starting from the World, the user can get

to all the data available in our database.

5. Frame 5: the data download information and meta-data description is shown here for geographic data (see Fig. 2). For grid data or image data both meta-data and sample images are displayed here (see Fig. 3).

4.3. Mapping tools

This site uses MapCafe to provide the mapping interface. MapCafe is a Java applet that comes with ArcView IMS. It provides a built-in setup and a default user interface similar to ArcView's, and a set of basic GIS functions. The interface provided by MapCafe is easy to manipulate if users are familiar with any desktop GIS system (ESRI White Paper, 1997a). The functions from the default set were used to customize the interface by setting the image size display on the web and making a few other modifications in Arc-

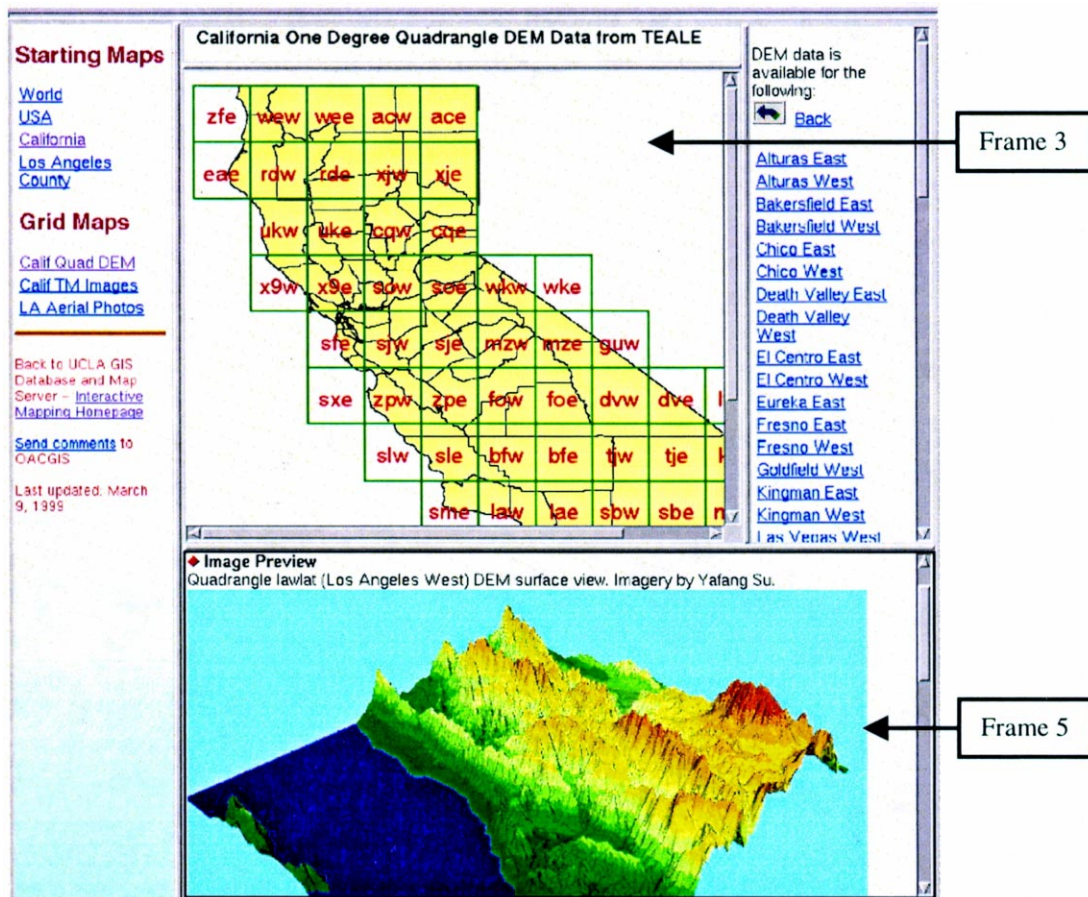


Fig. 3. Diagram of grid maps and data downloading example.

View's Avenue language and selecting among the MapCafe tools.

The standard MapCafe interface was provided with the customization of two tools of Identify and Hotlink. The Identify button allows users to view the information about the active layer in another browser window. The Hotlink button allows users to change the list of the available data on the frame at right according to the region selected.

4.4. Data sets for online mapping

The data sets that are available for manipulation online fall into the following two categories: maps needed for the visual interface and representative maps.

The maps used for the visual interface include the world country boundary map, the USA state boundary map, the California county boundary map and Los Angeles county maps because these are the focus areas in which we have the most data in the current UCLA GIS database. As the database is expanded, we add additional maps to the visual interface.

Representative (sample) maps are a small subset of the maps in our database. Each represents a different map category. As new geographic areas are added, the number of these maps will grow. We provide samples for each major database on the site.

Unauthorized users cannot download the data used for online mapping. The Identify button only identifies one or a few layers available with an online map, and each click only identifies one record. The print button only produces a JPEG image file which is of the displayed map not the actual data. The data license agreement is strictly maintained in this way. (ESRI White Paper, 1997b).

5. Conclusions

The UCLA GIS Database and Map Server web site not only delivers proprietary data to the UCLA community, but also makes GIS available for public access. In this way, any member of the UCLA community can easily access the maps and the GIS data available in our database. The site promotes GIS at UCLA. This site differs from a number of other sites provided by colleges and universities in that both proprietary data and interactive mapping are supported.

The UCLA solution is particularly designed for distributing proprietary data. The UCLA model may be of interest to universities and digital libraries that have proprietary data and that they would like to share only within their community.

References

- Baltazar, H., 1999. GIS comes down to earth. *PC Week* 16 (1), 14.
- Culpepper, R.B., 1998. Weave maps across the web. *GIS World* 11 (11), 46–52.
- Harder, C., 1998. *Serving Maps on the Internet*. Environmental Systems Research Institute Inc, Redlands, California, 130 pp.
- Limp, F., 1997. Weave maps across the web. *GIS World* 10 (9), 46–52.
- Plewe, B., 1997. *GIS Online: Information Retrieval, Mapping and the Internet*. OnWord Press, Albany, New York 311 pp.
- Zurier, S., 1996. *Geographic information systems: look around — you find GISes wherever feds are in Government*. *Computer News* 15 (12), 61.
- ESRI, 1999. ArcIMS version 3.0. *ArcUser*, July–September, 9.
- ESRI White Paper, 1997a. *The Future of GIS on the Internet*. ESRI, Redlands, California (11 pp).
- ESRI White Paper, 1997b. *How ESRI Uses the Internet/ Intranet to Deliver GIS On-line*. ESRI, Redlands, California (12 pp).

Internet references: selected data download sites

- [1] Alexandria Digital Library. <http://alexandria.sdc.ucsb.edu>.
- [2] BARD — San Francisco Bay Area Regional Database. <http://bard.wr.usgs.gov/index.html>.
- [3] Bay Area Shared Information Consortium. <http://badger-parl.com/>.
- [4] California State University, Northridge, Department of Geography. <http://geogdata.csun.edu>.
- [5] IESIN's SEDAC's Demographic Data Cartogram Service DDCarto. <http://plue.sedac.ciesin.org/plue/ddcarto/>.
- [6] Cornell University Geospatial Information Repository. <http://cugir.mannlib.cornell.edu>.
- [7] EROS Data Center Distributed Active Archive Center (EDC DAAC). <http://edcwww.cr.usgs.gov/landdaac/>.
- [8] ESRI ArcData Online. <http://www.esri.com/data/online>.
- [9] NOAA Server-access. <http://www.ncdc.noaa.gov/NOAAServer>.
- [10] TEALE Data Center. <http://www.gislab.teale.ca.gov>.
- [11] US Census Bureau. <http://www.census.gov/>.
- [12] UC Berkeley Earth Sciences & Map Library's GIS Home Page. <http://library.berkeley.edu/EART/digital/dem/dem.html>.
- [13] UK's National Spatial Data Sets Held at MIMAS. <http://www.kinds.ac.uk/kinds>.
- [14] USGS Geo-data and remotely sensed data. <http://mapping.usgs.gov>.
- [15] World Imagery Database by Microsoft TerraServer. <http://www.terra-server.com>.

Selected interactive GIS Mapping sites

- [16] A GIS Internet Map Server (in German). <http://xagisim-s.ag.ch/>.

- [17] ArcAustria IMS Demo. <http://www.datamed.co.at/arcview/imspage/aademo.html>.
- [18] Arizona State University. <http://gis1.inre.asu.edu>.
- [19] CARES Watershed Information Clearinghouse. <http://www.cares.missouri.edu/cwic>.
- [20] ESRI Live Demos. <http://maps.esri.com/ESRI/arcview/demos.htm>.
- [21] Florida Marine Research Institute/EPA Watershed Project. <http://www.fmri.usf.edu/surf>.
- [22] ForNet MapServer. <http://www.gis.umn.edu/fornet/docs/MapServer>.
- [23] MapInfo MapXtreme Demo Site. <http://www.mapinfo.com/software/mapxtreme/demo>.
- [24] MapQuest Maps and Directions. <http://www.mapquest.com/>.
- [25] Montana Maps Interactive. <http://orion.nris.state.mt.us>.
- [26] National Atlas of the United States. <http://www-atlas.usgs.gov/scripts/start.html>.
- [27] National Real Estate Finder. <http://www.realtor.com/>.
- [28] Nebraska Natural Resources Commission, Nebraska Map Interactive. <http://cavent.nrc.state.ne.us>.
- [29] NOAA NOS Office of Coast Survey. <http://anchor.ncd.noaa.gov/noaa/noaa.html>.
- [30] SanGIS Interactive Maps. <http://www.ruis.org/intmaps/>.