A VIRTUAL STRONG-MOTION DATA CENTER FOR DATA DISSEMINATION THROUGH THE INTERNET

R. Archuleta[1], A. Tumarkin[1], G. Lindley[1]
A. Shakal[2] and R. Borcherdt[3]

Institute for Crustal Studies
University of California at Santa Barbara[1]

California Strong-Motion Instrumentation Program
California Division of Mines and Geology[2]

United States Geological Survey[3]

ABSTRACT

As a possible contribution to the COSMOS mission to expand the use and application of strong-motion data, a virtual, Web-based data center is being created. The database will include data from COSMOS cooperating networks, which currently includes the California Strong Motion Instrumentation Program, the U.S. Geological Survey, the Army Corps of Engineers, and the U.S. Bureau of Reclamation. If accepted by COSMOS, the database would be managed under COSMOS.

The database has been created, and methods of access through the World Wide Web are under development. Currently, the database includes 740 three-component accelerogram records. The database structure consists of twelve tables, allowing for data about the accelerogram records, earthquakes, recording sites, instruments, networks, and station owners to be stored. The database also allows for all items in the database to have a comment or scientific reference attached to them.

The database has been created using Microsoft SQLServer7 and is running on a 450 MHz PC. Two mirror sites are planned. The database access software is being written in the programming languages Java and Perl. These languages were chosen so that the software will be portable across database and operating system platforms. Multiple Web access methods are being developed to allow searching of the database from earthquake, station, or accelerogram record parameters. Users will be able to select data from lists of earthquakes or stations, to query the database through Web forms pages, or to select data from dynamically created maps.

INTRODUCTION

COSMOS has been created with the mission to expand and modernize the acquisition and application of strong-motion data to increase public earthquake safety. In order to increase the ease of use and accessibility of strong-motion data, a data center is being created as a possible contribution to the COSMOS mission. The data center will include a strong-motion database that is accessible over the World-Wide Web, allowing the data to be downloaded over the Internet based on criteria supplied by the user. The Home Page is shown in Figure 1.

The features of the data center will include:

- True relational database with earthquake, station, instrument, network, and accelerogram parameters
- Data retrieved from agencies' existing FTP and Web sites
- Effective and efficient access for data users
- Effective feedback from users about the database, Web site, and the data itself
- Preservation of ownership and quality control for agencies collecting the data
- Appropriate credit for source networks

Users of the database will be able to access data based on a wide range of parameters, including peak ground acceleration, epicentral or hypocentral distance, earthquake or station name or location, site geology, housing structure, network name, station owner, or earthquake source parameters, such as magnitude, seismic moment, or focal mechanism. Multiple access methods will be made available to improve the effectiveness and efficiency for the database user.

In addition to ease of use, one of the advantages of the database will be the maintenance of the data by the agencies that collected the data. This will allow these agencies to maintain quality control over their data, to get appropriate credit for their work, helping to justify their missions, and it will give users an effective means of communication with these agencies.

This paper discusses the ongoing effort to meet these data center features. First, we discuss the details of the structure of the database. Then we describe the current and future data that is incorporated into the database. After that, we describe the hardware and software that is running the database. And finally, we discuss the methods of Web access that are currently being developed.

DESCRIPTION OF THE DATABASE

The database derives from an existing database produced at U.C. Santa Barbara, the Strong Motion DataBase. This database originally had four tables to hold information for the earthquakes, stations, accelerogram traces, and references or comments. In order to increase the functionality and store additional information, the new database has been expanded to twelve tables.

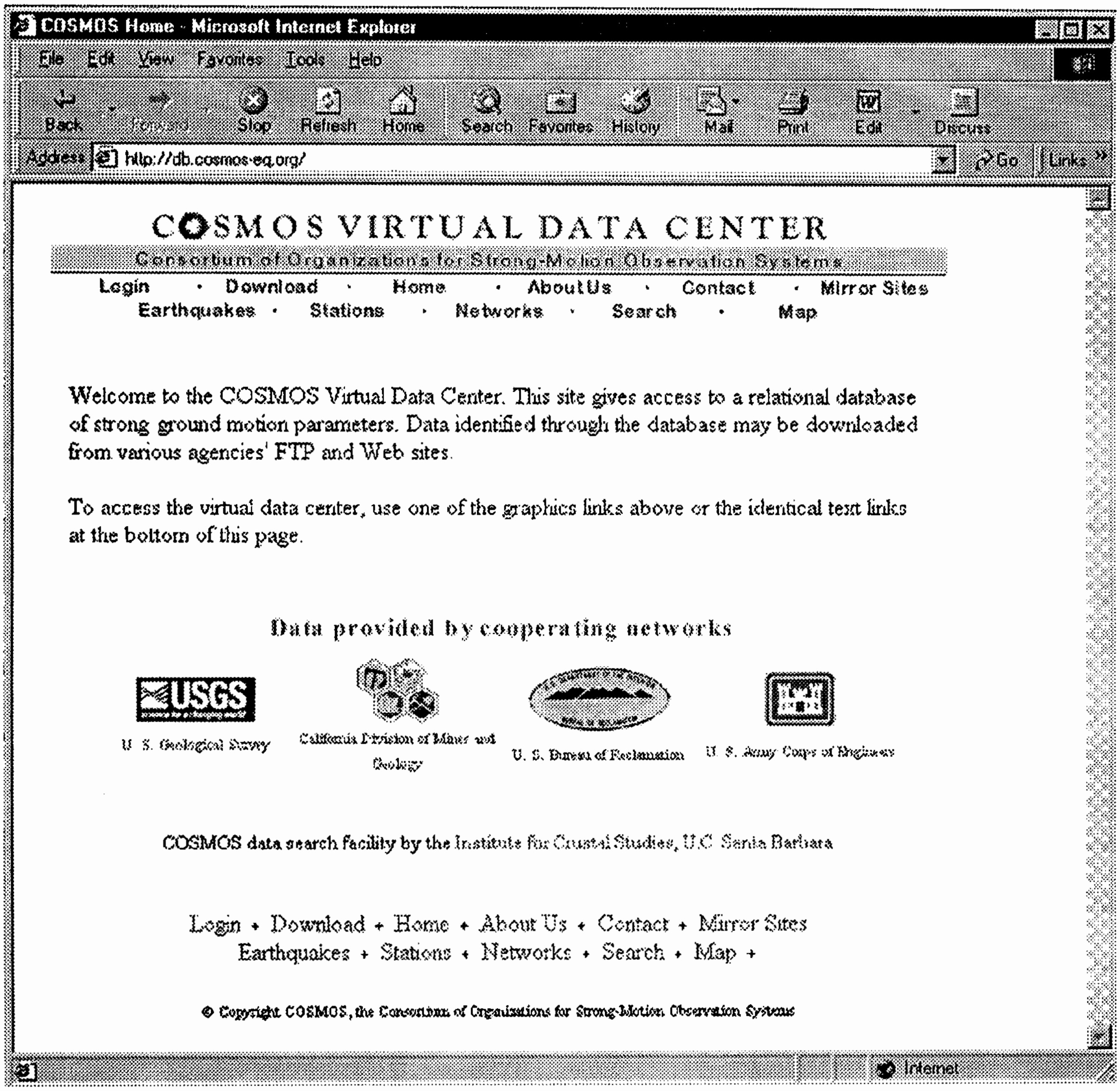

Figure 1. Proposed home page for the COSMOS Virtual Data Center.

By using twelve tables, the structure of the database more accurately matches the real-world structure of the data. The twelve tables are Station, Instrument, StationOwner, Network, Trace, Event, Region, Item, Download, WebUser, Comment, and CommentConnector. The relationships between the tables are shown in Figure 2. These tables can be placed into five groups: the station tables, the event tables, the trace or accelerogram table, the user tables, and the comment/reference tables.

The station tables consist of the Station, Instrument, StationOwner, and Network tables. The Station table is used to store characteristics of data recording locations. A station may be a single free-field instrument, or it may consist of multiple instruments such as might be found at an office building, a bridge overpass, or a dam. The Instrument table gives the location, type, and other parameters corresponding to the recording sensor. The StationOwner table contains information about the site's owner, including contact information for the user. The Network table identifies which network the data was derived from. These tables also include fields for Web addresses, where the user can find additional information about the stations, networks, and station owners.

The event tables consist of two tables, the Event and Region tables. The Event table contains information about the earthquakes, such as their location, magnitude, focal mechanism, and seismic moment. References for each of the parameters stored in the table can be included using the Comment and CommentConnector tables. The Region table increases the functionality of the database by allowing searches of the database that are limited to particular geographic regions.

The core of the database is the Trace table. This table contains the download addresses for the data, as well as information about the accelerogram traces such as the peak ground acceleration, epicentral distance, etc. The smallest "unit" of the database is a single accelerogram trace, allowing for maximum flexibility in use of the database. However, through the Instrument and Station tables, a user can easily select data from a particular location. For example, a user interested in data from the 1989 Loma Prieta, California earthquake might select all of the data recorded at Anderson Dam, or just the data from the downstream record, or just the vertical recording from the downstream record.

Comments, references, and additional Web or FTP addresses will be stored in the Comment and CommentConnector tables. These tables have been structured for maximum efficiency, allowing each element of the database to be commented or referenced multiple times if necessary. The same reference or comment can also apply to multiple elements within the database. This is important so that the user can track down the sources of information in the database. For example, to say that a particular earthquake had a surface magnitude of 6.9 is not very useful unless a reference is attached to it.

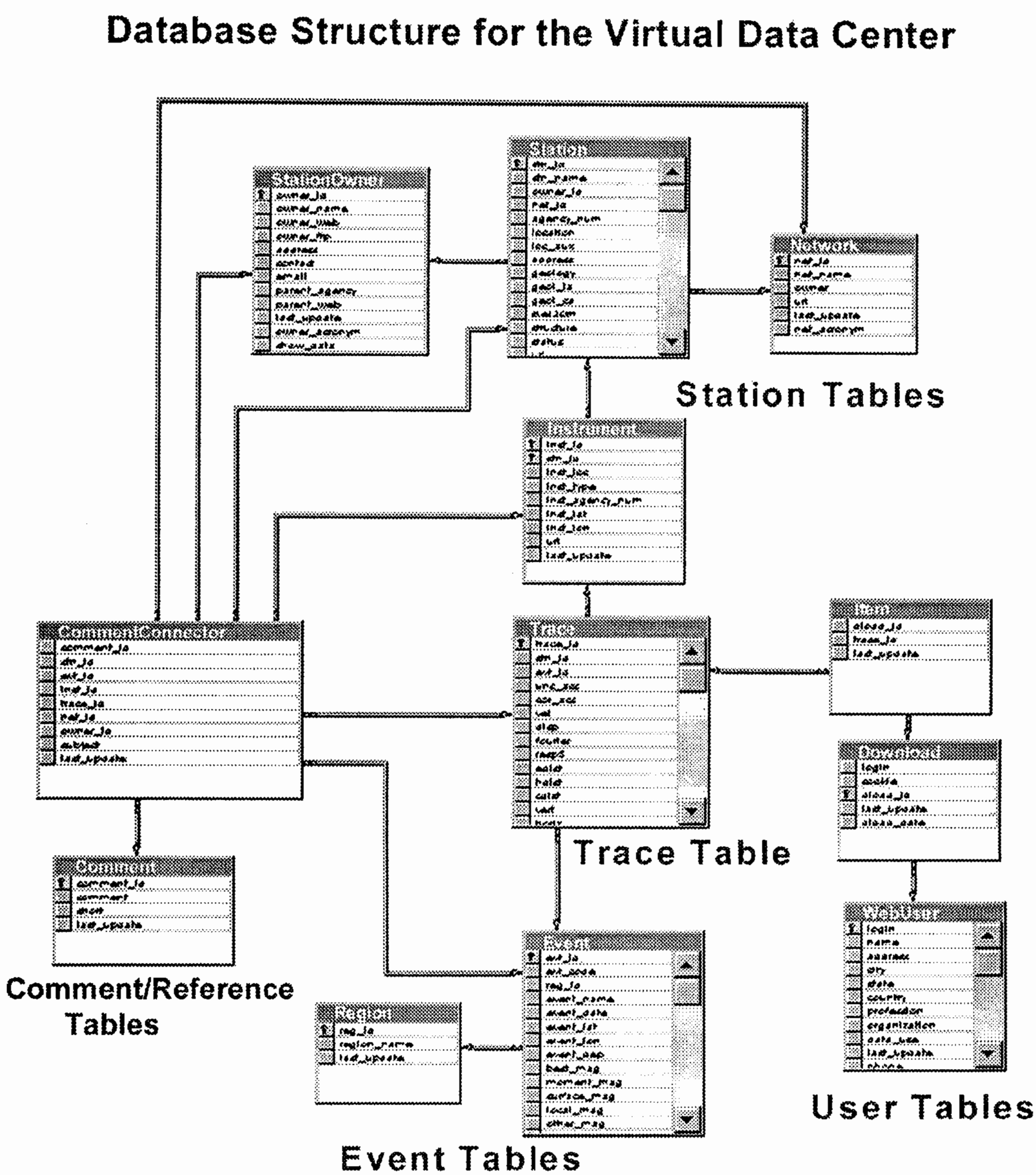

Figure 2. Basic structure of the database, showing the relationships between the various tables.

Finally, there are three tables that keep track of user information in the database: the WebUser, Download, and Item tables. Most of the information stored in these tables will be supplied on a voluntary basis, and will allow the agencies involved in the project to assess what types of uses the data is being put to. This will allow these agencies to more accurately justify their efforts and to adjust those efforts to meet the needs of the end-users. These tables will also allow the Web site to use a "shopping cart" type model, whereby the users will add data to their "cart" and download all of the data when they are through searching the database.

CURRENT AND FUTURE DATA

The database has already been populated with data from the USGS, CDMG, ACOE, and USBR networks, as well as the older Caltech "blue book" data. The database contains 740 3-component records recorded at 561 stations from 104 earthquakes. Ninety-two earthquakes are from California; five are from the Pacific Northwest; four are from the central and eastern United States; three are from Alaska, and one is from Mexico. The oldest event in the database is the 1933 Long Beach, California earthquake. The most recent is the 1996 Duval, Washington earthquake. The events in the database are summarized in Table 1.

Methods are being developed to add data from future earthquakes in a timely fashion. The cooperating agencies will use an automated method to add data to the database. In addition, as more networks become contributing COSMOS members, their data will be added to the database. The data in the database can be expected to grow rapidly.

HARDWARE/SOFTWARE

The hardware and software for the site were chosen for their ability to handle a moderate-sized database in an efficient and cost-effective manner. The database software is Microsoft SQLServer7. It is installed on a Microsoft Windows NT platform running a 450 MHz Intel chip. It is desirable, however, that the database be easily portable to other database software or to a different operating system. In order to meet this goal, the software that will drive the database on the Web site is being written in the languages Java and Perl. These languages are not highly dependent on a particular operating system or software manufacturer. The other option under consideration, Microsoft Active Server Pages, has not been chosen as it would limit future portability of the database.

In addition to the main server, which will operate, at least initially, from UC Santa Barbara, two mirror sites are planned, one to be operated by the US Geological Survey and the other to be operated by the California Division of Mines and Geology. This will reduce downtime of the database and will split the load among three different locations.

Table 1. Events in the Virtual Data Center

LONG BEACH CA	1933-03-11, 01:54:07
SOUTHERN CALIFORNIA	1933-10-02, 09:10:17
BAJA CALIFORNIA MEXICO	1934-12-30, 13:52:00
HELENA MONTANA	1935-10-31, 11:38:00
IMPERIAL VALLEY CA	1938-04-12, 19:29:00
IMPERIAL VALLEY CA	1938-06-06, 02:42:00
IMPERIAL VALLEY CA	1938-06-06, 12:35:00
EL CENTRO CA	1940-05-19, 04:36:40
EL CENTRO CA	1940-05-19, 04:41:02
EL CENTRO CA	1940-05-19, 04:42:13
SANTA BARBARA CA	1941-07-01, 07:50:54
TORRANCE-GARDENA CA	1941-11-14, 08:41:36
BORREGO VALLEY CA	1942-10-21, 16:22:13
WESTERN WASHINGTON	1949-04-13, 11:56:00
IMPERIAL VALLEY CA	1951-01-24, 07:17:02
KERN COUNTY CA	1952-07-21, 11:52:14
KERN COUNTY CA	1952-07-21, 12:05:00
KERN COUNTY CA	1952-07-21, 17:42:44
KERN COUNTY CA	1952-07-21, 19:41:22
SAN LUIS OBISPO CA	1952-11-22, 07:46:00
IMPERIAL VALLEY CA	1953-06-14, 04:17:29
WHEELER RIDGE CA	1954-01-12, 23:33:49
IMPERIAL COUNTY CA	1955-12-17, 05:17:21
IMPERIAL COUNTY CA	1955-12-17, 06:07:29
PORT HUENEME CA	1957-03-18, 18:56:00
SAN FRANCISCO CA	1957-03-22, 11:44:00
PUGET SOUND WA	1965-04-29, 07:28:00
SOUTHERN CALIFORNIA	1965-07-16, 07:46:00
PARKFIELD CA	1966-06-28, 04:26:00
BORREGO MOUNTAIN CA	1968-04-09, 02:28:59
LYTLE CREEK CA	1970-09-12, 14:30:52
SAN FERNANDO CA	1971-02-09, 14:00:41
SAN FERNANDO CA	1971-02-09, 14:01:33
SAN FERNANDO CA	1971-02-09, 14:01:50
SAN FERNANDO CA	1971-02-09, 14:02:24
SAN FERNANDO CA	1971-02-09, 14:03:25
SAN FERNANDO CA	1971-02-09, 14:04:34
SAN FERNANDO CA	1971-02-09, 14:05:50
CENTRAL CALIFORNIA	1971-03-08, 23:08:07
ADAK ALASKA	1971-05-02, 00:00:00
POINT MUGU CA	1973-02-21, 14:45:57
IMPERIAL VALLEY CA	1974-12-06, 12:13:08
ANCHORAGE ALASKA	1975-01-01, 00:00:00
IMPERIAL VALLEY CA	1975-01-23, 17:02:29
IMPERIAL VALLEY CA	1975-06-20, 05:48:24
HORSE CANYON CA	1975-08-02, 00:14:07
IMPERIAL VALLEY CA	1976-11-04, 05:48:20
IMPERIAL VALLEY CA	1977-10-20, 10:29:35
IMPERIAL VALLEY CA	1977-10-21, 13:24:24
IMPERIAL VALLEY CA	1977-10-30, 05:30:14
IMPERIAL VALLEY CA	1977-11-14, 00:11:35
IMPERIAL VALLEY CA	1977-11-14, 02:05:48
IMPERIAL VALLEY CA	1977-11-14, 05:36:55
IMPERIAL VALLEY CA	1977-11-14, 12:20:20
COYOTE DAM CA	1978-03-26, 00:27:00
SANTA BARBARA CA	1978-08-13, 22:54:53
SOUTHEASTERN ALASKA	1979-02-28, 00:00:00

COYOTE LAKE CA	1979-08-06, 17:05:22
IMPERIAL VALLEY CA	1979-10-15, 23:16:53
IMPERIAL VALLEY CA	1979-10-16, 06:58:32
LIVERMORE CA	1980-01-24, 19:00:09
LIVERMORE CA	1980-01-27, 02:33:35
ANZA CA	1980-02-25, 10:47:38
MAMMOTH LAKES CA	1980-05-25, 16:33:44
MAMMOTH LAKES CA	1980-05-25, 16:49:27
MAMMOTH LAKES CA	1980-05-25, 19:44:51
MAMMOTH LAKES CA	1980-05-25, 20:35:48
MAMMOTH LAKES CA	1980-05-26, 18:57:56
MAMMOTH LAKES CA	1980-05-27, 14:50:57
WESTMORLAND CA	1981-04-26, 12:09:28
NEW HAMPSHIRE USA	1982-01-19, 00:14:42
ANZA CA	1982-06-15, 23:49:21
ENOLA ARKANSAS	1982-06-26, 00:00:00
ENOLA ARKANSAS	1982-07-05, 00:00:00
COALINGA CA	1983-05-02, 23:42:38
COALINGA CA	1983-05-09, 02:49:11
COALINGA CA	1983-06-11, 03:09:52
COALINGA CA	1983-07-09, 07:40:51
COALINGA CA	1983-07-22, 02:39:54
COALINGA CA	1983-07-22, 03:43:01
COALINGA CA	1983-07-25, 22:31:39
COALINGA CA	1983-09-09, 09:16:13
MORGAN HILL CA	1984-04-24, 21:15:18
REDLANDS CA	1985-10-02, 23:44:12
NORTH PALM SPRINGS CA	1986-07-08, 09:20:44
WHITTIER NARROWS CA	1987-10-01, 14:42:20
WHITTIER CA	1987-10-04, 10:59:38
ELMORE RANCH CA	1987-11-24, 01:54:14
SUPERSTITION HILLS CA	1987-11-24, 13:15:56
LOMA PRIETA CA	1989-10-18, 00:04:15
UPLAND CA	1990-02-28, 23:43:36
SIERRA MADRE CA	1991-06-28, 14:43:54
JOSHUA TREE CA	1992-04-23, 04:50:00
CAPE MENDOCINO-PETROLIA CA	1992-04-25, 18:06:04
CAPE MENDOCINO-PETROLIA CA	1992-04-26, 07:41:00
CAPE MENDOCINO-PETROLIA CA	1992-04-26, 11:18:00
LANDERS CA	1992-06-28, 11:57:34
BIG BEAR CA	1992-06-28, 15:05:36
SCOTTS MILL OREGON	1993-03-25, 00:00:00
NORTHRIDGE CA	1994-01-17, 12:30:55
EUREKA CA	1994-09-01, 15:15:52
SOUTH LAKE TAHOE CA	1994-09-12, 12:23:42
PARKFIELD CA	1994-12-20, 10:27:47
DUVAL WA	1996-05-03, 00:00:00

Table 1. Events in the Virtual Data Center continued.

WEB SITE LAYOUT AND FUNCTIONALITY

Four basic methods of Web access are being developed for the database. These methods will allow the user to select data by viewing a list of earthquakes in the database, by viewing a list of stations in the database, by user input onto HTML forms pages, or through clickable maps showing earthquake and station locations. In addition to the dynamic database access pages, static Web pages will give basic information about the data center to the user including contact information and information about the mirror sites. See Figure 3 for a basic layout of the Web site.

Users will download data via a "shopping cart" model, which is commonly seen on Web sites today. As the user browses through the database, they will be given the opportunity to add the data they are interested in to a download bin. When the user is ready, they can select an option to download the data they have previously added to their bin. The downloaded data will be grouped by agency as a means of giving additional attribution to the contributing agencies.

Each page of the Web site will contain links to the database access methods, to the static pages of the Web site, and to the data download pages. A proposed home page showing these links can be seen in Figure 1.

To make the Web site as efficient as possible, the responses to database queries will be organized into a three-tiered structure. The three tiers will be station, instrument (3-component record), and trace. The user will have the option of selecting all of the data from a station (which may have multiple instruments), from a particular instrument, or selecting just a single accelerogram trace.

CONCLUSIONS

A relational, Web-accessible database of parameters related to strong-motion data is being developed for possible inclusion in a COSMOS virtual data center. The database will allow users to selectively acquire strong-motion data corresponding to the user's exact needs. The database will contain information about the accelerogram records and related information about the recording sites, networks, and earthquakes. The data itself will be downloaded from existing Web and FTP sites. The database will have the capacity to add references or comments to every element of the database. Multiple Web access methods will be made available as well as a means for the users to contact the agencies that recorded the data. The database will improve access to strong-motion data, and will provide an important way for the agencies that collected strong-motion data to gather information about the uses their data is being put to.

Web Site Layout for the Virtual Data Center

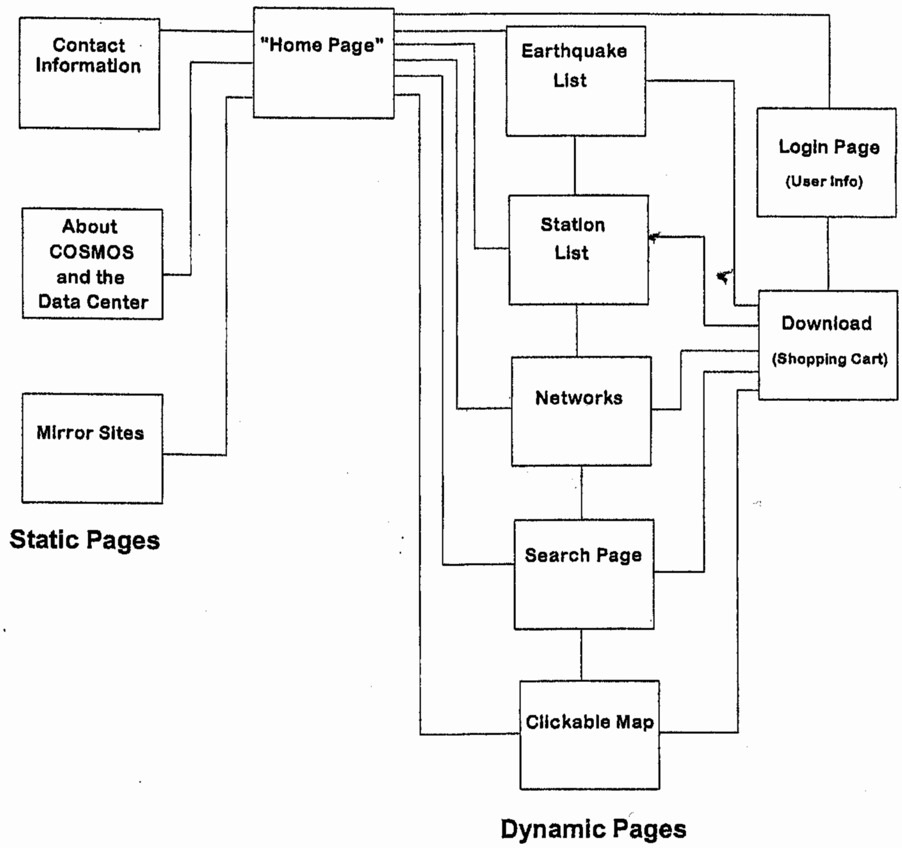

Figure 3. Layout of the Web site that will access the database.