GALILEO: Behind the Scenes

Brad Baxter
Information Analyst, GALILEO Development
University of Georgia
Main Library
Athens, Georgia 30602
bmb@mail.libs.uga.edu

GALILEO: Georgia Library Learning Online
Office of Information and Instructional Technology
Board of Regents of the University System of Georgia
1865 West Broad Street
Athens, Georgia 30606-3539
www.galileo.peachnet.edu

GALILEO, Georgia Library Learning Online, is the product of the state of Georgia initiative, A Vision for One Statewide Library. It is a World Wide Web based database server offering access for participating institutions to over 100 databases indexing thousands of periodicals and scholarly journals. GALILEO was born out of the desire to take advantage of the most current technology to advance Georgia libraries' services to patrons. Creating GALILEO required the widespread cooperation of librarians and technical staff. GALILEO takes advantage of the World Wide Web and the proliferation of web browsers to enhance both face-to-face instruction and distance learning. This paper will briefly describe some of the “behind the scenes” aspects of the current GALILEO system. These include the people who govern and develop the system, the networks that make it possible and the organization of its server computers. Also covered is the software that makes up GALILEO as well as a brief look at the locally maintained databases. By knowing more about what happens in the background, users of GALILEO may better understand both the power and some of the limitations of the system.
GALILEO: Behind the Scenes

GALILEO, Georgia Library Learning Online, is the product of the state of Georgia initiative, *A Vision for One Statewide Library: GALILEO* (USG, 1995). The GALILEO home page is accessible on the World Wide Web at [http://www.galileo.peachnet.edu](http://www.galileo.peachnet.edu). The home page provides a link, About GALILEO, which leads to a series of papers that describe how the system came into being. Also at this site are answers to frequently asked questions (FAQ) and information about the GALILEO/PeachNet Service Center.

**Background**

GALILEO is a World Wide Web based resource that combines several services:

- Information about the 34 University System of Georgia (USG) institutions through a link to the USG web site
- Library catalogs for the USG institutions and other institutions in the state
- Suggested Internet resources to help new patrons readily find useful information at other web sites
- A rich assortment of both locally maintained and remote databases

There are more than twenty databases that are available to all citizens of Georgia and more than 100 that are available to educational institutions who pay for the extended access. These databases provide access to citations and abstracts from thousands of periodicals and scholarly journals and the full text of articles from over 2000 journal titles. The wealth of resources in GALILEO enhances both face-to-face instruction and distance learning.

Jayne Williams, of the Board of Regents Office of Information and Instructional Technology, is the GALILEO Project Director. She states in *The Long, but Short History of GALILEO* (1997), “GALILEO was born out of a history of necessity melding with ideas, ideas melding with opportunities, and opportunities melding with resources.” That paper describes the chronology of events that led to the official launching of GALILEO on September 21, 1995.

By taking advantage of the most current technology, GALILEO is able to advance Georgia Libraries’ services to patrons. These patrons include not only USG students, faculty and staff, but individuals at all the educational institutions across the state, including K-12 schools, and every citizen of Georgia who uses a public library. Creating GALILEO required the widespread cooperation of librarians and technical staff. It also required of everyone a certain courage to believe they could make such a new and complex project a success.

**Overview**

GALILEO makes use of the World Wide Web and the proliferation of web browsers. The emerging phenomenon of the web simplifies developing an interface for GALILEO, because the client/server infrastructure is already in place. That is to say, everyone can readily obtain a copy of the web client software, i.e., a web browser, and the GALILEO team has ready access to the latest web server programs. With this “hard part” taken care of, the GALILEO developers have only to design the look and feel and the navigability of the user interface.

This paper and its corresponding presentation will briefly describe some of the behind the scenes aspects of the current GALILEO system. By knowing more about what happens in the background, users of GALILEO may better understand both the power and some of the limitations of the system.

The slides that accompany the presentation are available at this address: [http://www.galileo2.peachnet.edu/public/GALILEO001behindthescenes.html](http://www.galileo2.peachnet.edu/public/GALILEO001behindthescenes.html)

Online versions of this paper are available at these addresses:

(HTML) [http://www.galileo2.peachnet.edu/public/GALILEObehindthescenes.html](http://www.galileo2.peachnet.edu/public/GALILEObehindthescenes.html)

(PDF) [http://www.galileo2.peachnet.edu/public/GALILEObehindthescenes.pdf](http://www.galileo2.peachnet.edu/public/GALILEObehindthescenes.pdf)
People

It goes without saying (but can never be said enough) that the success of any undertaking depends on the efforts of individuals. An undertaking as large as GALILEO requires the combined efforts of many individuals and groups. A number of committees, charged with responsibility for the system, help to coordinate the work of everyone involved.

RACL, the Regent's Academic Committee on Libraries, began in 1968 and consists primarily of the USG university library directors. RACL was instrumental in forwarding the GALILEO initiatives and continues to play a leading role in its ongoing development. As outlined in GALILEO Governance Structure (RACL, 1997), both the GALILEO Advisory Council and Steering Committee include, by design, members of RACL.

Figure 1: People

Figure 1 illustrates the relationships among some of the key committees and groups responsible for the governance and development of GALILEO.

The GALILEO Advisory Council consults with the Steering Committee on strategic direction and budget review, and advises on external evaluation procedures for the project. The GALILEO Steering Committee manages the project.

OIIT, the Office of Information and Instructional Technology, serves as the contracting authority for GALILEO, implements and manages the PeachNet network, and houses the GALILEO/PeachNet Service Center. Also, in conjunction with Georgia State University (GSU) and the University of Georgia (UGA), OIIT maintains a staff of computer programmers and the key computer equipment and software for GALILEO.

The Steering Committee appoints subcommittees as necessary, and two such committees are the Reference Subcommittee and the Technical Operations Subcommittee. The Reference Subcommittee works with the computer programmers and service center staff to develop a comprehensive, functional and effective user interface. The Technical Operations Subcommittee reviews technical questions and develops and reviews implementation schedules. Other subcommittees include the User Group Subcommittee and the Assessment Subcommittee.

Other individuals contributing to GALILEO’s success are the numerous reference librarians who take time every day helping and training patrons who use GALILEO—in addition to providing all of the other services that libraries offer. Finally, the technical support staff at every educational institution deserve much appreciation. Their efforts ensure that the local networks and workstation computers function well.
Networks

Simplistically speaking, a network is a number of computers connected by wires. These “wires” actually may be a wide variety of materials. They can range from wireless configurations to fiber optic cables to coaxial cables to twisted pair copper wires. The phrase “connected by” lumps together an incredible array of routers, bridges, switches and hubs, and a very complex interaction among many different network software programs.

- A LAN is a Local Area Network. An institution or a department within an institution may maintain a LAN, which may connect as few as two or as many as hundreds of computers.

- PeachNet is Georgia’s statewide educational network. It existed long before GALILEO, and is vital to GALILEO's success. PeachNet connects all of Georgia's educational institutions’ networks together.

- The Internet is the worldwide network. A multitude of networks across the globe—including PeachNet—are interconnected via the Internet. The World Wide Web (WWW) is a subset of the Internet.

Computers connected to the Internet interact using various communication “languages” or protocols. The WWW is that set of computers that interact using HTTP, the Hypertext Transfer Protocol. Other protocols include Telnet, FTP, Gopher, Z39.50 and many more.

As figure 2 illustrates, a large institution may have many local area networks. Each LAN normally has one central server computer. These servers connect to the campus backbone, a network itself that interconnects all the LAN's and provides (typically) one or more T1 connections to PeachNet. A T1 connection has a capacity of 1.544 megabits per second. A small institution might have a single LAN (probably in the library) whose server computer connects to PeachNet. A single user with an MCI Campus account may connect to PeachNet via modem. PeachNet serves not only to interconnect the educational institutions in Georgia, but also is a conduit via BBN Planet and UUNET to the rest of the Internet and the World Wide Web. BBN Planet and UUNET are Internet Service Providers (ISP) contracted by PeachNet for this service.
GALILEO is fortunate to be able to take advantage of state of the art computer systems. These server computers are larger and more powerful than those available just a few years ago. The interaction between these computers is very complex, and enables GALILEO to perform high-capacity transmission services very quickly.

The two brands of servers, IBM and Sun Microsystems, are both UNIX-based. The IBM computers run the AIX operating system, and the Sun computers run the Solaris system. The servers are located at Georgia State University, GSU, and the University of Georgia, UGA. Each site maintains several computers, and these service the bulk of database searching on GALILEO. A process called load balancing distributes the user load between these two sites, and between computers at the sites. Balancing the load assures that the server machines provide the best possible response time.

Figure 3 shows a diagram of the server computers at GSU and UGA. The primary GALILEO address, http://www.galileo.peachnet.edu, points to the switch server. The web server program running on the switch server presents the GALILEO home page. The switch server is a Sun Ultra I with one 170 megahertz central processing unit (CPU)—the main computer chip—and 64 megabytes of random access memory (RAM.)

When a user clicks DATABASES on the home page, the load balancing process takes place, and the switch server connects the user to one of the three primary GALILEO servers: Server A, Server M or Server N. Server A is a Sun SPARCserver-1000 with six 60 megahertz CPU’s and 640 megabytes of RAM. Server M is a Sun Ultra II with two 200 megahertz CPU’s and 1 gigabyte (1,000 megabytes) of RAM. Server N is identical to Server M.

Server A, Server M and Server N are “user session” servers each running SiteSearch’s Webz web server. Webz (pronounced “web-zee”) is a special type of web server that keeps track of user state information. This means that the computer remembers what each user does, so the software services requests in the context of what the user has done before. See the discussion of software below for more information about SiteSearch.

Server B and Server C at GSU and Servers E, F, G and H at UGA are database servers. When a patron chooses one of the local databases and performs a search, that query may pass to any of the database server computers. Queries from Server M and Server N always pass to one of the UGA database server computers (Server E, Server F, etc.) Queries from Server A pass to Server B or Server C, unless the database is not on those computers. Databases such as EBSCO and Georgia Government Publications are only at UGA, and all queries for these databases pass to one of
the UGA database server computers. By separating user session processing overhead from local database searching overhead, a sort of division of labor takes place that allows the entire group of computers together to handle an ever-increasing user load.

Server B is a Sun Ultra II with two 200 megahertz CPU's and 512 megabytes of RAM. Server C is identical to Server B. Servers E, F, G and H are four nodes of an IBM SP/2 computer. Each node is an IBM RS/6000 with one 66 megahertz CPU and 256 megabytes of RAM.

When a user submits a search for a FirstSearch database such as WorldCat, that query passes (via the Internet) to a database server at OCLC FirstSearch in Ohio. When a user chooses to search one of the remote database interfaces such as Cambridge Scientific or GaleNet, the user moves to that site and communicates directly with that vendor's remote user interface.
Software

GALILEO uses standard web software: web server programs and web client programs. Web servers are programs that run on server computers and communicate using the HTTP protocol. When a user connects to the server computer, the web server program reads web pages from the computer hard disk and transmits them to the user.

A web client, more commonly called a web browser, is a program that runs on a user's computer. When a user clicks on a link, the web browser transmits that request to a web server across the net. When the web browser receives the web page back, it translates the internal format (HTML, Hypertext Markup Language) into a formatted document.

GALILEO also uses SiteSearch, an advanced web-based database searching system. GALILEO purchased SiteSearch from OCLC, Online Computer Library Center, Inc., http://www.oclc.org. SiteSearch includes programs that use the HTTP and Z39.50 protocols and a proprietary search engine, Newton. These programs combine into a software system tailored to GALILEO's specific needs. Z39.50 is the Information Retrieval Service and Protocol standard. This protocol allows database searching and retrieval software to share widely varying bibliographic data in a standard way across networks.

OCLC continues to improve SiteSearch and is looking forward to introducing a new version written almost entirely in the Java programming language. Once in place, the new version should provide an even more powerful searching environment for GALILEO patrons.

![Figure 4: Software](image)

Figure 4 diagrams the software programs that serve patrons. Apache (internally called “httpd”) is the web server program that runs on the switch server, and is the primary entry point to GALILEO. Apache presents the GALILEO home page. When a user clicks DATABASES the “Switch” program randomly selects from the three primary user session servers (Server A, Server M or Server N), and makes a connection to the SiteSearch Webz server (httpman). The Webz server “spawns” a GALILEO session (httpgate) for each user. Each user session occupies some of that computer’s memory (RAM) for as long as that session is active.

For a local database, the SiteSearch software allows a direct search between a user session and the Newton database search engine. However, because the user session servers are separate from the database servers, GALILEO uses Z39.50 searches for local databases. These searches create local Z39.50 sessions on the database servers. These local sessions pass queries to the database engine and format results to send back to the user session for display.

Almost all of the OCLC FirstSearch databases are also available via Z39.50 searches. These search requests create remote Z39.50 sessions on the FirstSearch database server computers. These sessions in turn are the go-between from user sessions to the remote FirstSearch database search engines.
When a user chooses a remote database such as Academic Press or Encyclopædia Britannica, the Webz server program passes the user to the remote vendor's web server program. This new web server takes over completely and communicates directly to the user (via the user's web browser). This completely bypasses both the Webz server and the user's GALILEO session. While the user is communicating to the remote system, the “abandoned” GALILEO session will stay in place for fifteen minutes, after which it will time out, terminate and free up the memory it was occupying.
Databases

In addition to the many databases available from remote vendors like OCLC FirstSearch and Cambridge Scientific, GALILEO maintains several local databases. Many of these local databases contain the full text of journal articles. Maintaining local databases is efficient, economical and allows greater flexibility than relying exclusively on remote databases.

- Local databases are efficient because the data travels a shorter route from the server computer to the user.
- Local databases are economical, because a particular license is for the whole database. This avoids the per-search fees that some vendors charge.
- Local databases allow greater flexibility because programs can modify the data and indexing when updating the database.

Some of the local databases, such as Georgia Government Publications and Georgia Libraries Journal List, are products of the GALILEO project, and users from anywhere in the world may search them.

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Size in Megabytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodical Abstracts</td>
<td>13,779</td>
</tr>
<tr>
<td>ABI/Inform</td>
<td>11,164</td>
</tr>
<tr>
<td>Business Dateline</td>
<td>5,696</td>
</tr>
<tr>
<td>Newspaper Abstracts</td>
<td>6,954</td>
</tr>
<tr>
<td>Dissertation Abstracts</td>
<td>5,798</td>
</tr>
<tr>
<td>Current Contents</td>
<td>19,858</td>
</tr>
<tr>
<td>EBSCO Index Complete (3 years)</td>
<td>20,877</td>
</tr>
<tr>
<td>ERIC (Educational Resources Information Center)</td>
<td>3,903</td>
</tr>
<tr>
<td>Georgia Government Publications (records)</td>
<td>40</td>
</tr>
<tr>
<td>Georgia Government Publications (images)</td>
<td>6,485</td>
</tr>
<tr>
<td>Georgia Libraries Journal List</td>
<td>397</td>
</tr>
<tr>
<td>Total:</td>
<td>94,951</td>
</tr>
<tr>
<td></td>
<td>(95 Gigabytes)</td>
</tr>
</tbody>
</table>

Figure 5: Databases

Figure 5 is a table of all the local GALILEO databases showing the size of each and the total size all together. A megabyte is approximately 1,000 kilobytes. A kilobyte is approximately 1,000 individual letters or numbers, making a megabyte approximately 1,000,000 characters. A gigabyte is approximately 1,000 megabytes. Therefore, the 95 gigabytes of data in all the local databases amount to ninety-five thousand million characters of information.

Conclusion

GALILEO is complex, powerful and rich. The complex computer and software systems take advantage of emerging technology. The computers and programs are fast and powerful, transmitting many megabytes of data in seconds. With access to the Internet, and by providing large local databases, a rich bounty of information is in reach.

Through the close cooperation of many people and groups, GALILEO has evolved into its current configuration. It will, of course, continue to evolve.
References

