Is This a Geolibrary? A Case of the Idaho Geospatial Data Center

Maria Anna Jankowska
and Piotr Jankowski

The article presents the Idaho Geospatial Data Center (IGDC), a digital library of public-domain geographic data for the state of Idaho. The design and implementation of IGDC are introduced as part of the larger context of a geolibrary model. The article presents methodology and tools used to build IGDC with the focus on a geolibrary map browser. The use of IGDC is evaluated from the perspective of access and demand for geographic data. Finally, the article offers recommendations for future development of geospatial data centers.

In the era of integrated transnational economies, demand for fast and easy access to information has become one of the great challenges faced by the traditional repositories of information—libraries. Globalization and the growth of market-based economies have brought about, faster than ever before, acquisition and dissemination of data, and the increasing demand for open access to information, unrestricted by time and location. These demands are mobilizing libraries to adopt digital information technologies and create new methods of cataloging, storing, and disseminating information in digital formats.

Libraries encounter new challenges constantly. Participation in the global information infrastructure requires them to support public demand for new information services, to help the society in the process of self-education, and to promote the Internet as a tool for sharing information. These tasks are becoming easier to accomplish thanks to the growing number of digital libraries. Since 1994, when the Digital Library Initiative originated as part of the National Information Infrastructure Program, the Internet has accommodated many digital libraries with spatial data content. For example, the Electronic Environmental Library Project at the University of California, Berkeley (http://elib.cs.berkeley.edu/) provides botanical and geographic data; the University of Michigan Digital Library Teaching and Learning Project (www.si.umich.edu/UMDL/) focuses on earth and space sciences; the Carnegie Mellon's Informedia Digital Video Library (www.informedia.cs.cmu.edu) distributes digital video, audio, and images

with text; and the Alexandria Digital Library at Santa Barbara (http://alexandria.sdc.ucsb.edu/) provides geographically referenced information. The Alexandria Digital Library is of special interest in this article because it implements a model of a geolibrary. A geolibrary stores georeferenced information searchable by geographic location in addition to traditional searching methods such as by author, title, and subject.

The purpose of this article is to present the Idaho Geospatial Data Center (IGDC) in the larger context of a geolibrary model. IGDC is a digital library of public-domain geographic and statistical data for the state of Idaho. The article discusses methodology and tools used to build IGDC and contrast its capabilities with a geolibrary model. The usage of IGDC is evaluated from the perspective of access and demand for geographic data. Finally, the article offers recommendations for future development of geospatial data centers.

Geographic Information Systems for Public Services

Terms such as digital, electronic, virtual, or image libraries have existed long enough to inspire diverse interpretations. The broad definition by Covi and King concentrates on the main objective of digital libraries, which is the collection of electronic resources and services for the delivery of materials in different formats.[1] The common motivation for initiatives leading to the development of digital libraries is to allow conventional libraries to move beyond their traditional roles of gathering, selecting, organizing, accessing, and preserving information. Digital libraries provide new tools allowing their users not only to access the existing data but also to create new information. The creation of new information using the existing data sources is essential to the very idea of the digital library. Since the information in a digital library exists in virtual form, it can be manipulated instantaneously by computer-based information processing tools. This is not possible using traditional information media (e.g., paper, microfilm) where the information must first be transferred from non-digital into digital format.

Since late 1994, when the U.S. National Science Foundation founded the Alexandria Digital Library Project, the number of Internet sites devoted to spatially referenced information has grown dramatically. Today, it would require a serious expenditure of time and effort to visit all geographic data sites created by state agencies, universities, and commercial organizations. In 1997 Karl Musser wrote, "there are now more than 140 sites featuring interactive maps, most of which have been created in the last two years."[2] This incredible boom in publishing

Maria Anna Jankowska (majanko@uidaho.edu) is Associate Network Resources Librarian, University of Idaho Library, and **Piotr Jankowski** (piotrj@uidaho.edu) is Associate Professor, Department of Geography, University of Idaho, Moscow, Idaho.

spatial data is possible thanks to geographic information system (GIS) technology and data development efforts brought about by the rapidly increasing use of GIS. This new technology provides its users with capabilities to automate, search, query, manage, and analyze geographic data using the methods of spatial analysis supported by data visualization.

Traditionally, geographic data were presented on maps considered as public assets. According to a Norwegian survey, the aggregate benefit accrued from using maps was three times the total cost of their production, even though maps provided only static information.[3] Today, the conventional distribution of geographic data on printed maps has become less efficient than distributing them in the digital format through wide area data networks. This happened largely due to GIS's ability to separate data storage from data presentation. As a result, data can be presented in a dynamic way, according to users' needs. Often GIS is termed "data mixing system" because it can process data from different sources and formats such as vector-format maps with full topological and attribute information, digital images of scanned maps and photos, satellite data, video data, text data, tabular data, and databases.[4] All of these data types provide a rich informational infrastructure about locations and properties of entities and phenomena distributed in terrestrial and subterrestrial space.

The definition of GIS changes according to the discipline using it. GIS can be used as a map-making machine, a 3-D visualization tool, and as an analytical, planning, collaboration, and business information management tool. Today, it is hard to find a planning agency, city engineering department, or utility company (not to mention individual Internet users) that has not used digital maps. This is why the number of users seeking spatial data in digital format has increased so dramatically. Data discovery can be for GIS users the most time-consuming part of using the technology.[5] As a result, libraries are faced with the growing demand for services that help discover, retrieve, and manipulate spatial data. The Web greatly improved the availability and accessibility of spatial data but, at the same time, stimulated public interest in using geographic information.

The continuing migration to popular operating systems (i.e., Microsoft Windows family) and the adoption of their common functionality has brought GIS software to many desktops. Tools such as ArcView GIS from Environmental Systems Research Institute, Inc. (ESRI, www.esri.com) or MapInfo from MapInfo Corporation (MapInfo, www.mapinfo.com) have become popular GIS desktop systems. New software tools such as ArcExplorer, released by ESRI, are focused on making GIS more accessible, simpler, and available for use by the public. By taking advantage of the popularity of the Web,

attempts are being made to gain a wider acceptance of GIS. In the wake of the simplification of GIS tools and improved access to spatial data, a new exciting area of GIS use has recently emerged—public participation GIS.[6] Public participation GIS by definition is a pluralistic, inclusive, and nondiscriminatory tool that focuses on the possibility of reducing the marginalization of societies by means of introducing geographic information operable on a local level.[7] It promotes an understanding of spatial problems by those who are most likely to be affected by the implementation of problem solutions, and encourages transfer of control and knowledge to these parties. This approach leads to a broader use of GIS tools and spatial data and creates new challenges for libraries storing and serving geographic data in digital formats. Broadening the use of data and GIS tools requires attention to data access. Traditional libraries have often fulfilled the crucial role of being an impartial information provider for all parties involved in public decision-making processes. Will they be capable of serving the society in this capacity in the digital age?

Geolibrary as a Repository of Georeferenced Information

According to Brandon Plewe, the user of spatial data can choose among seven types of distributed geographic information services available on the Internet.[8] They range from raw data download, through static map display, metadata search, dynamic map browsing, data processing, Web-based GIS query and analysis, to net-savvy GIS software. Yet, another important new category of geographic data service that can be added to this list is geolibrary.

Goodchild defines a geolibrary as a library filled with georeferenced information where the primary basis of representation and retrieval are spatial footprints that determine the location by geographic coordinates. "The footprints can be precise, when they refer to areas with precise boundaries, or they can be fuzzy when the limits of the area are unclear."[9] According to Buttenfield, "the value of a geolibrary is that catalogs and other indexing tools can be used to attach explicit locational information to implicit or fuzzy requests, and once accomplished, can provide links to specific books, maps, photographs, and other materials."[10] A geolibrary is distinguished from a traditional library in being fully electronic, with digital tools to access digital catalogs and indexes. It is anticipated that most of the information is archived in digital form. The value of a geolibrary is that it can be more than a traditional, physical library in electronic form.[11]

Since its introduction, the concept of a geolibrary has been synonymous with the Alexandria Digital Library (ADL) project. Once ADL was defined as the Internet-based archive providing comprehensive browsing and retrieval services for maps, images, and spatial information.[12] A more recent definition characterizes ADL as

> a geolibrary where a primary attribute of collection objects is their location on Earth, represented by geographic footprints. A footprint is the latitude and longitude values that represent a point, a bounding box, a linear feature, or a complete polygonal boundary.[13]

According to Goodchild (1998) a geolibrary's components include:

- **The browser**—a specialized software application running on the user's computer and providing access to geolibrary via a computer network.
- **The basemap**—a geographic frame of reference for the browser's searches. A basemap provides the image of an area corresponding to the geographical extent of geolibrary collection. For the worldwide collection this would be the image of the Earth. For the statewide collection this could be the image of a state. The basemap may be potentially large, in which case it is more advantageous to include it in the browser then to download it from a geolibrary server each time a geolibrary is accessed.
- **The gazetteer**—the index that links place names to a map. The gazetteer allows geographic searches by place name instead of by area.
- **Server catalogs**—collection catalogs maintained on distributed computer servers. The servers can be accessed over a network with the browser, utilizing basic server-client architecture.

The value of a geolibrary lies in providing open access to a multitude of information with geographic footprints regardless of the storage media. Because all information in a digital library is stored using the same digital medium, traditional problems of physical storage, accessibility, portability, and concurrent use (e.g., many patrons wanting to view the one and only copy of a map) do not exist.

▌ Idaho Geospatial Data Center

In 1996, inspired by the ADL project, a team of geographers, geologists, and librarians started to work on a digital library of public-domain geographic data for the state of Idaho. The main goal of the project was the development of a geographic digital data repository accessible through a flexible browsing tool. The project was funded by a grant from the Idaho Board of Education's Technology Incentive Program. The project resulted in the creation of the Idaho Geospatial Data Center (IGDC, http://geolibrary.uidaho.edu). The first in the state of Idaho, this digital library is comprised of a database containing geospatial datasets, and GeoLibrary software that facilitates access, browsing, and retrieval of data in popular GIS data formats including Digital Line Graph (DLG), Digital Raster Graphics (DRG), USGS Digital Elevation Model (DEM), and U.S. Bureau of Census TIGER boundary files for the state of Idaho. The site also provides an interactive visual analysis of selected demographic/economic data for Idaho counties. Additionally, the site provides interactive links to other Idaho and national spatial data repositories. The key component of the library is the GeoLibrary software. The name "GeoLibrary" is not synonymous with the model of geolibrary defined by Goodchild (1998). It was rather adopted as a reference to a geolibrary browser—one of the components of the geolibrary.

The GeoLibrary browser (GL) supports online retrieval of spatial information related to the state of Idaho. It was implemented using Microsoft Visual Basic 5.0/6.0 and ESRI MapObjects technology. The software allows users to query an area of interest using a search based on map selection, as well as selection by area name (based on USGS 7.5-minute quad naming convention). Queries return GIS data including DEMs, DLGs, DRGs, and TIGER files. Queries are intended both for professionals seeking GIS-format data and nonprofessionals seeking topographic reference maps in the DRG format.

The interface of GL consists of three panels resembling the Microsoft Outlook user interface. Our intent in designing the interface was to have panels that would be used in the following order. First, the map panel is used to explore the geographic coverage of the geolibrary and to select the area of interest. Next, the query panel is used to execute a query, and finally the result panel allows the user to analyze results and to download spatial data. Users can use a shortcut to go directly to the query panel and type their query. Both approaches result in the output being displayed as the list of files available for download from participating servers.

The map panel (figure 1) includes a navigable map of Idaho, a vertical command toolbar, and a map finder tool. The command toolbar allows the user to zoom in, zoom out, pan the map, identify by name the entities visible on the map canvas, and select a geographic area of interest. Geographic entity name identification was implemented as a dynamic feature whereby the name of entity changes as the user moves the mouse over the map. Spatial selection provides a tool to select a rectangular area of interest directly on the map canvas. The map finder provides additional means to simplify the exploration of the map.
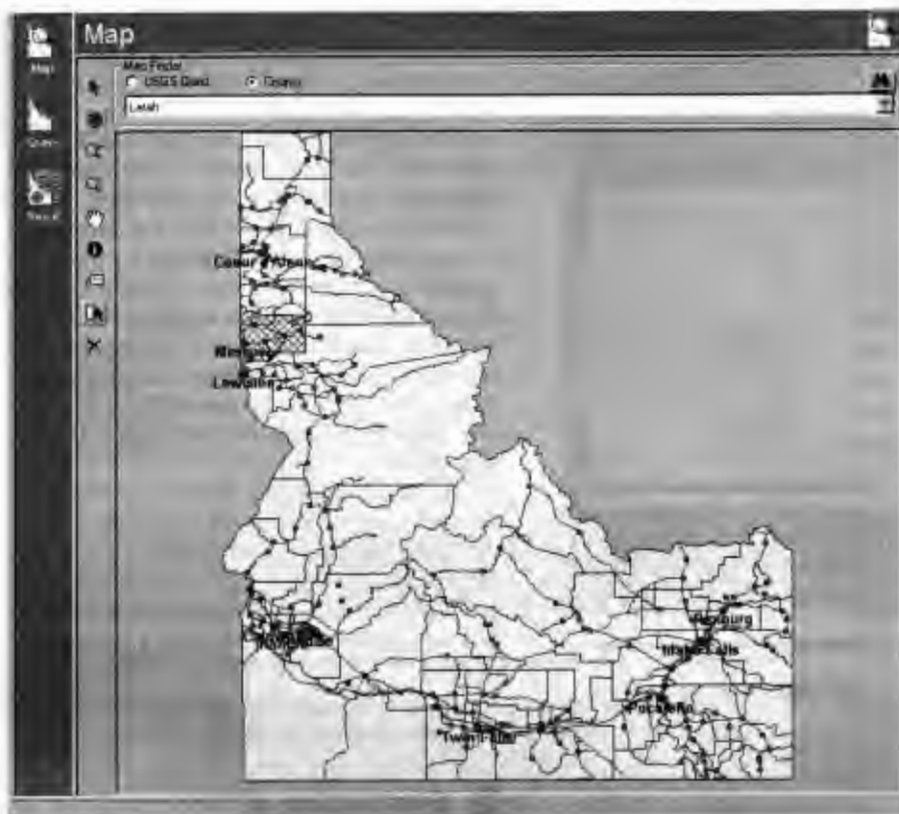
Figure 1. Map panel. The vertical toolbar provides zooming, panning, as well as labeling and simple feature querying capabilities. The map finder allows finding and selecting an area by county or USGS quad name. The screen copy here presents the selection of Latah County in Idaho.

The user can select a county or a quad name and zoom in on the selected geographic unit.

The query panel (figure 2) allows the user to perform a query, based either on the selection made on the map or a new selection using one of the available query tools (figure 3). In the latter case, the user can enter geographic coordinates (in decimal degrees) defining the area of interest. This approach is equivalent to selecting a rectangular area directly on the map, and will return all data files that spatially intersect with the selected area. Optionally, the user can handpick quads of interest from the list. Finally, a name can be entered to execute a more flexible query. For instance, the search containing the word "Moscow" returns spatial data related to three quads containing "Moscow" within their names. The query is executed when the user presses the Query button. After the results are received, the application automatically switches to the results panel.

The results panel shows the outcome of the query and includes important information about the data files: their size, type, projection, scale, the name of the server providing the data, as well as the access path (figure 4). Based on this information, the user has the option of manually connecting to the server, using FTP protocol, and retrieving the selected files. A much more convenient approach, however, is to rely on GL software to automatically retrieve the files through the software interface. As an option, the result of the query can also be exported to a plain HTML document that contains links to all listed files. This feature can be very useful in the case of multi-file files selected by the user and slow or limited-time Internet access. This way the user can open the saved list of files in a Web browser and download individual files as needed, without having to download all the files at once and tie up the Internet connection for a long period of time.

The result panel provides a flexible way to review and organize the outcomes of queries before commencing the download. One can sort files by name, size, scale, projection, and server name. This feature may be useful if the user decides to retrieve data of only one type (e.g., DEMs), of one scale, or when the user prefers to connect only to a specific sever. In addition, individual records as well as entire file types can be selected to prevent files from being downloaded. The user can also remove selected files to scale down the set of data in the list.

One of the most important assets of the GL browser is that all of the user activities described up to this point, with the exception of file download, take place entirely on the client-side without any network traffic. In fact, area/file selection as well as queries do not require an active Internet connection. Map exploration is based on vector-format maps contained in GL software and queries are run against the local database. Such an approach limits bandwidth consumption and unnecessary network traffic. Internet connection is only necessary to perform retrieval of selected files.
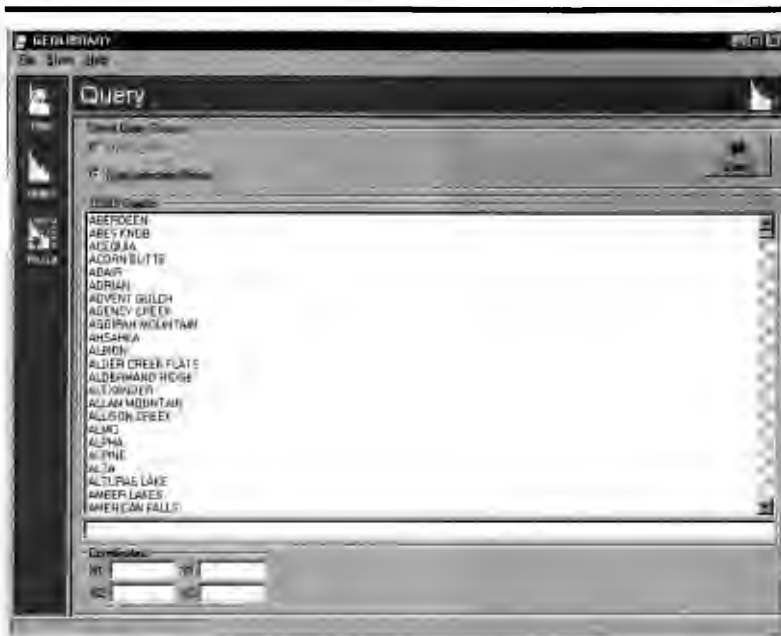
Figure 2. Query panel. The interface was set to query spatial selection from the map panel.
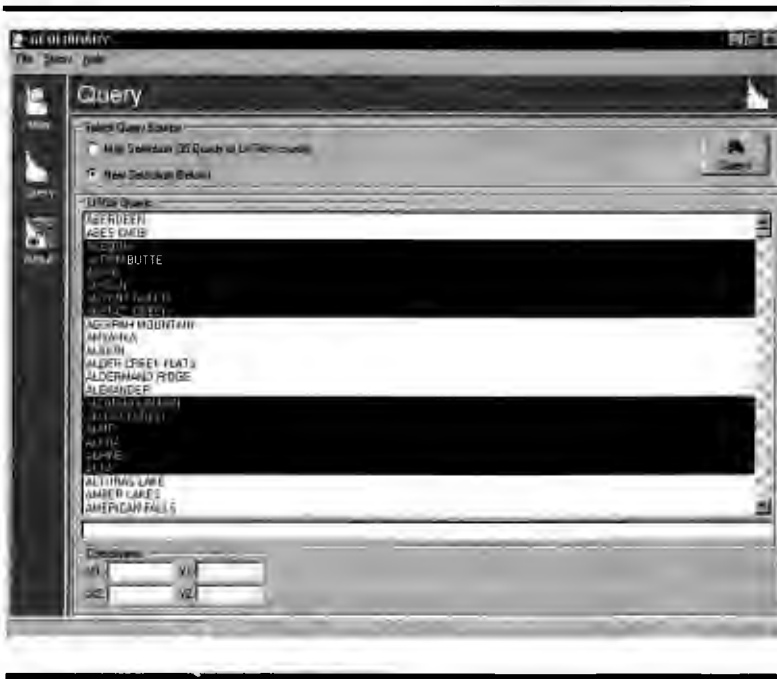

Figure 3. Query panel. The query is based on the selection of USGS quads. Optionally, the user can enter geographic coordinates of the area or a text to search.

The vulnerability of the client-side approach to data query is to be left with a potentially outdated local database. In order to prevent this problem from happening, the GL is equipped with a database synchronization mechanism that allows users to keep up with the server database updates. The client-side database, contained in GL software, which mirrors the schema of the server database, can be synchronized automatically or by the user's request. In either case, the GL client contacts the server-based database synchronizer on the server side and handles all necessary processes. Since the synchronization is limited to database record updates, the network traffic is kept low, making GL suitable for limited Internet connections.

IGDC is an open solution. New local datasets can be added or removed making the collection easily adaptable to different geographical areas. In addition, datasets can physically reside on multiple servers, taking full advantage of the Internet's distributed nature.

Evaluation of IGDC Use

Geospatial information is among the most common public information needs; almost 80 percent of all information is geographic in nature. Published research reflecting those needs and the role of libraries in resolving them is not extensive. The efforts of federal, state, and local agencies collecting digital geospatial data and the growth of GIS created an interest in the role of libraries as repositories of geospatial data.[14]

The main obstacle to providing access to digital spatial information is its complexity. This is why the user-friendly interface is critical for presenting spatially referenced information.[15] The IGDC has been a first attempt at creating a user-friendly interface in the form of a map-based data browser allowing the users to access and retrieve geographic datasets about Idaho. In order to track and evaluate the use of geospatial data, WebTrends software was installed on the IGDC server. The WebTrends software produces customized Web log statistics and allows tracking information on traffic and
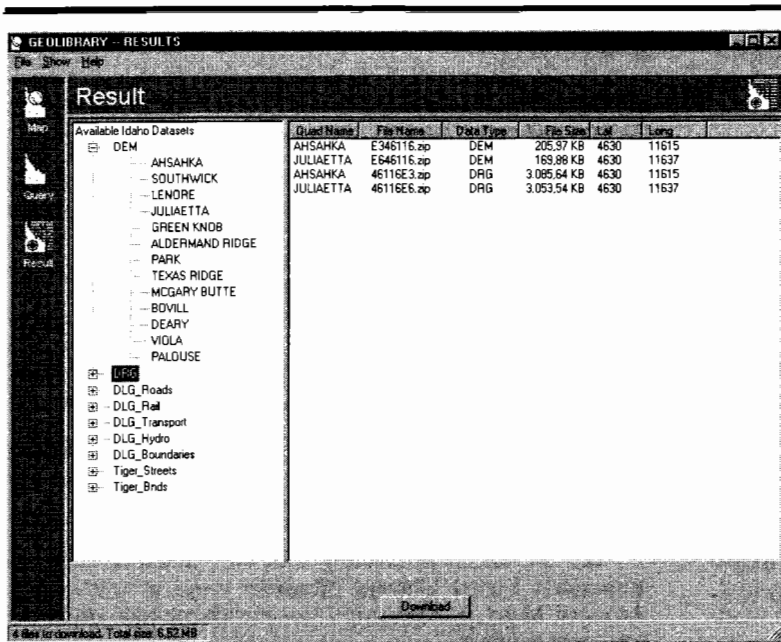
Figure 4. The results panel. Results of a query can be sorted; individual items can be removed from the list or can be deselected to prevent them from being downloaded.

datasets dissemination. During a one-year timeframe the number of successful hits was more than twenty-five thousand. Almost 40 percent of users came from .com domain, 35 percent were .net domain users, 15 percent were .org, and 10 percent were .edu users (figure 5).

Tracking the geographic origin of users by state, the biggest number of users came from Virginia, followed by Washington, California, Ohio, and Idaho. The high number of users from Virginia can be explained by the linking of the IGDC site to one of the most popular geospatial data sites in the country—the United States Geological Survey (USGS) site. Eighty-four percent of user sessions were from the United States; the rest originated from Sweden, Canada, and Germany. The average number of hits per day on weekdays was around one hundred customers. The most popular retrievable information were Digital Raster Graphics (DRG) data that present scanned images of USGS standard series topographic maps at 1:24,000 scale. Digital Elevation Models (DEM) and Digital Line Graphs (DLG) were less popular. The Tiger boundary files for the state of Idaho were in small demand. The popularity of DRG-format maps and the fact that most of the users accessed IGDC via the USGS Web site makes plausible a speculation that most of the users were non-GIS specialists interested in general reference geographic information about Idaho including topography and basic land use information.

Since the opening of IGDC for public use (April 1998), the GeoLibrary map browser was downloaded 1,352 times. The software proved to be relatively easy to use by the public. Out of forty-four bug reports/user questions submitted to IGDC, most were concerned with filling out the software registration form and not with software failure.

The IGDC project spurred an interest in geographic information among students, faculty, and librarians at the University of Idaho. In a direct response to this interest, the University of Idaho library installed a new dedicated computer at the reference desk with GeoLibrary software to access, view, and retrieve IGDC data.

■ Conclusion

Idaho Geospatial Data Center is the first geospatial digital library for the state of Idaho. It does not fulfill all requirements of a geolibrary model proposed by Goodchild and others. The IGDC has only two components of the geolibrary model; they are the GeoLibrary map browser and the basemap. The main difference between the GeoLibrary map browser and a Web-based browser solution adopted by other spatial repositories is a client-side solution to geospatial data query and selection. Spatial data query is done locally on the user's machine, using the library data base schema contained in the GeoLibrary map browser. This saves time by eliminating client-server communication delays during data searches, gives the user an experience of almost instantaneous response to queries, and reduces the network communication to the data download time.

In comparison with the geolibrary model, IGDC is missing the gazetteer. This component can help improve the ease of user navigation through a geospatial data collection. The other useful component includes online mapping and spatial data visualization services. The idea of such services is to provide the user with a simple-to-operate mapping tool for visualizing and exploring the results of user-run queries. One such service, currently under implementation at IGDC, includes thematic mapping of economic and demographic variables for Idaho using Descartes software.[16] Descartes is a knowledge-based system supporting users in the design and utilization of thematic maps. The knowledge base incorporates domain-independent visualization rules determining which map presentation technique to employ in response to the user selection of variables. An intelligent

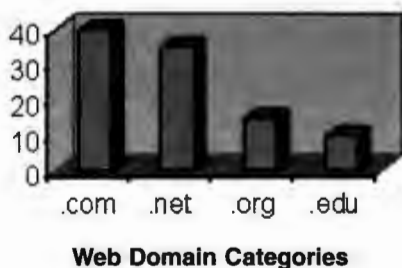**Distribution of IGDC Users (in %) by Domain**

Figure 5. Distribution of IGDC Users in Percent by Origin Domain

map generator such as Descartes can enhance the utility of a geolibrary by providing tools to transform georeferenced data into information.

References and Notes

1. L. Covi and R. King, "Organizational Dimensions of Effective Digital Library Use: Closed Rational and Open Natural Systems Models," *Journal of the American Society for Information Science* 47, no. 9 (1996): 697.

2. K. Musser, "Interactive Mapping on the World Wide Web." (1997) Accessed March 6, 2000, www.min.net/~boggan/mapping/thesis.htm.

3. T. Bernhardsen, *Geographic Information Systems* (Arendal, Norway: Viak IT and Norwegian Mapping Authority, 1992), 2.

4. Ibid., 4.

5. J. Stone, "Stocking Your GIS Data Library," *Issues in Science and Technology Librarianship.* (Winter 1999). Accessed March 6, 2000, www.library.ucsb.edu/istl/99-winter/article1.html.

6. P. Schroeder, "GIS in Public Participation Settings." (1997.) Accessed June 2, 1999, www.spatial.maine.edu/ucgis/testproc/schroeder/ucgisdft.htm.

7. W. J. Craig and others, "Empowerment, Marginalization, and Public Participation GIS," Report of a Specialist Meeting Held under the Auspices of the Varenius Project. Santa Barbara, California, Oct. 15–17, 1998, NCGIA, UC Santa Barbara.

8. B. Plewe, *GIS Online: Information Retrieval, Mapping, and the Internet* (Santa Fe, N.M.: OnWord Pr., 1997), 71–91.

9. M. F. Goodchild, "The Geolibrary," in *Innovations in GIS 5: Selected Papers from the Fifth National Conference on GIS Research UK (GISRUK),* ed. S. Carver. (London: Taylor and Francis, 1998), 59. Accessed March 6, 2000, www.geog.ucsb.edu/~good/Geolibrary.html.

10. B. P. Buttenfield, "Making the Case for Distributed GeoLibraries." (1998) Accessed March 6, 2000, www.nap.edu/html/geolibraries/app_b.html.

11. Ibid.

12. M. Rock, "Monitoring User Navigation through the Alexandria Digital Library," (master's thesis abstract, 1998). Accessed March 6, 2000, http://greenwich.colorado.edu/projects/rockm.htm.

13. L. L. Hill and others, "Geographic Names the Implementation of a Gazetteer in a Georeferenced Digital Library. *D-Lib Magazine* 5, no. 1 (1999). Accessed March 6, 2000, www.dlib.org/dlib/january99/hill/01hill.html.

14. M. Gluck and others, "Public Librarians' Views of the Public's Geospatial Information Needs," *Library Quarterly* 66, no. 4 (1996): 409.

15. B. P. Buttenfield, "User Evaluation for the Alexandria Digital Library project." (1995) Accessed March 6, 2000, http://edfu.lis.uiuc.edu/allerton/95/s2/buttenfield.html.

16. G. Andrienko and others, "Thematic Mapping in the Internet: Exploring Census Data with Descartes," in Proceedings of TeleGeo '99, First International Workshop on Telegeoprocessing, Lyon, May 6–7, R. Laurini, ed. (Seiten, France: Claude Bernard Univ. of Lyon, 1999), 138–45.