The Alabama Metadata Portal: http://portal.gsa.state.al.us

By Philip T. Patterson

Geological Survey of Alabama
420 Hackberry Lane
P.O. Box 869999
Tuscaloosa, AL 35468-6999
Telephone: (205) 247-3611
Fax: (205) 349-2861
e-mail: PPatterson@gsa.state.al.us

INTRODUCTION

In recent years federal, state, and local government entities in Alabama have made substantial investments in the collection, management, and use of geospatial data. However, there has been no large scale effort to share data effectively and efficiently. The result was unnecessary expenditures in redundant data creation.

Most Alabama Geographic Information Systems (GIS) users currently have broadband internet access. The increased network connectivity and high data-transmission rates have produced the expectation that large amounts of data can be accessed instantly. This demand for data access has motivated the Alabama Emergency Management Agency (AEMA) and the Geological Survey of Alabama (GSA) to collaborate in developing the geospatial data portal, which allows cooperators and users to search for, discover, and access geospatial data (GSA, 2006).

BACKGROUND

Before starting the project, extensive research on a variety of data delivery options was performed. The majority of the options were related to data clearinghouses, which are mainly useful for specific types of static data like imagery, civic boundaries, center lines, etc. However, the data delivery website to be built would not be intended for static data alone. The need was to build a robust compilation of all different types of vector and raster data, ranging from general datasets to obscure data specific to individual projects. Also long-term administration responsibilities for this type of complex compilation site were a concern for GSA. Eventually, the grant for site development would end, and GSA would have to support managing and updating the site from internal resources.

With support from Environmental Systems Research Institute Inc. (ESRI), we addressed this concern with a modified out-of-the-box application using open-source web applications in conjunction with ArcIMS, ArcSDE, and an underlying database management system (DMS). The resulting site provides the functions of a clearinghouse for general data and a search engine for unique data. It also offers semi-automated administration, which allows users, as well the administrator, to manage the site. This solution is ideal in addressing the data delivery goals and the long-term administration concerns posed by this project.

CONNECTION

This search engine and download site provide the framework for a mutual geospatial user community of organizations and stakeholders that facilitates discovery, sharing, and delivery of GIS content and services. The portal also facilitates the organization of content and services such as directories, search tools, community information, and support resources applications.

The underlying structure of the portal is a three-part generalized connection as follows (Figure 1): (1) the portal connects to a data provider's metadata library, which grants users the rights to publish specified metadata records to the portal's online catalog; (2) the data user connects to the portal's search option to locate data using the portal's search engine without physically browsing through the stakeholder's data; and (3) the data users will connect to the data provider for download, data captures, or the identification of the data resource.

By storing only metadata records in our catalog, we have the ability to index a large amount of virtual data, and more importantly, the GSA and AEMA will not have to store the physical data. Our goal is to automate the tasks of data discovery and distribution so that once portal connections are complete, minimal maintenance is required from the hosting agency.

ARCHITECTURE

A portal is essentially a master web site, which is connected to a web server and contains a database of metadata information about geographic data and services. The services are exposed as web applications using open

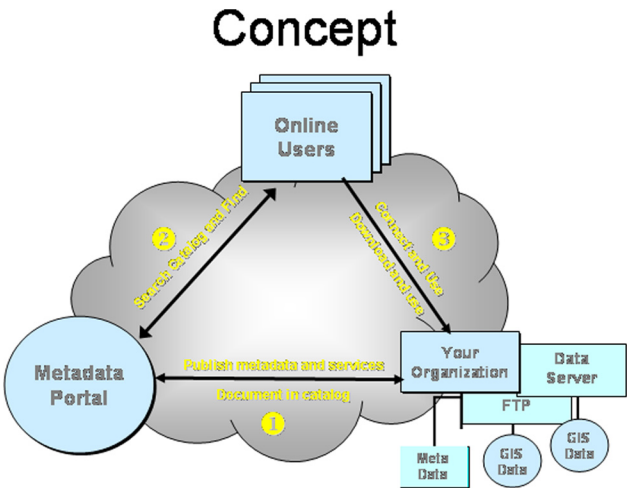

Figure 1. Generalized data partnership and user connection concept (modified from ESRI, 2004).

source environments (Tomcat, Java, html, http, xslt, xml, and jsp) to provide a user-friendly and visually appealing interface.

The architecture of the metadata server, which connects to all indexed metadata records, relies on three existing ESRI products: ArcIMS, ArcGIS, and ArcSDE. The ArcIMS provides the framework and architecture on which the metadata server runs. The ArcGIS ArcCatalog application serves as an authoring and publishing tool. The ArcSDE stores published metadata in records inside a relational database (ESRI, 2004). ArcIMS introduces a new approach to serving map products over the internet through a Java-based application management environment that includes mapping services and map design tools to support a variety of internet map services (ESRI, 2004). Main components associated with the ArcIMS communication architecture and web applications are identified in Figure 2.

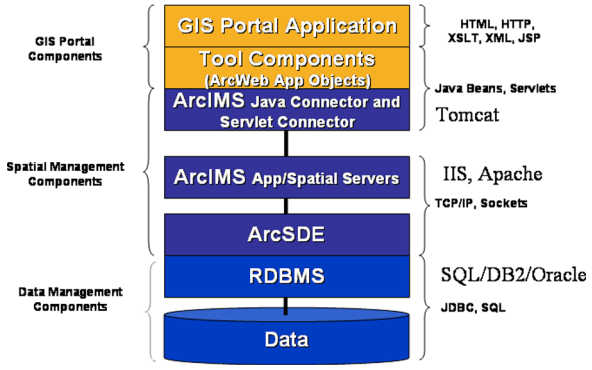

Figure 2. GIS software environment (Modified from ESRI, 2004).

THE PORTAL'S ONLINE INTERFACE COMPONENTS

Home Page

The home page shown in Figure 3 is the access point for all online components, and it provides quick access to the most popular data applications. From the home page, a user can do a basic keyword search, navigate to the map viewer, find help information, and access the quick links to downloadable data, GIS projects and services, and GIS resources.

The home page is also where users login to their accounts. A user account is not necessary to access the portal, but it increases user capability and enhances functionality. There are five distinct user levels of the portal based on a top-down hierarchy; that is, higher level users can do everything a lower level user can do. From lowest to highest, these include:

1. Anonymous users can be anyone. These users have the ability to browse the site and use three basic

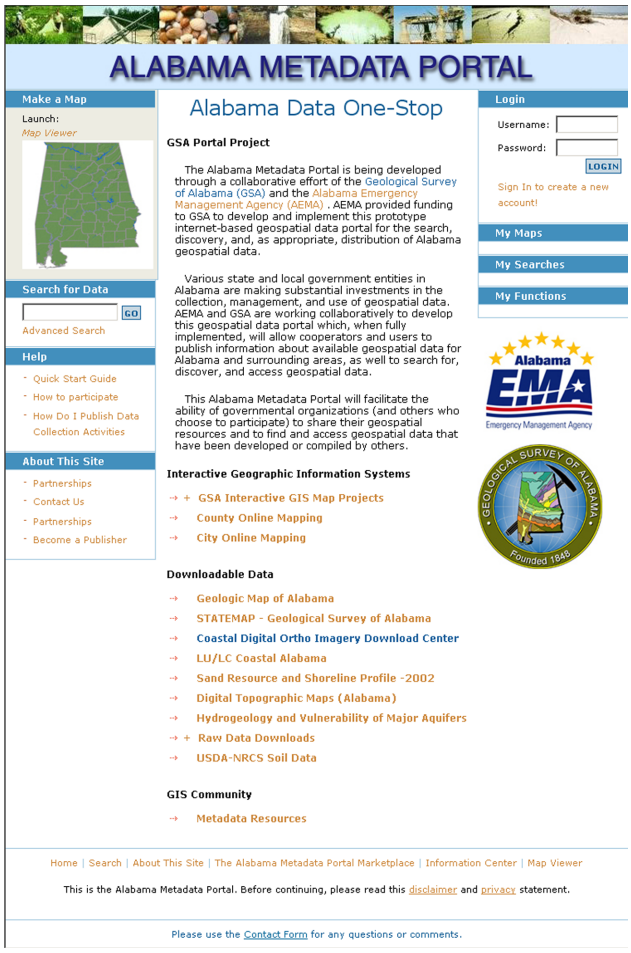

Figure 3. Example of the portal homepage.

online components: home page, map viewer, and search page.
2. Public Users have the ability to save their created maps from the map viewer and save their data searches, which will be available on the users' home page.
3. Publisher Users have the ability to create, publish, and manage their metadata online.
4. Channel Managers have the ability to create and publish a quick link on the home page.
5. Administrators check metadata for accuracy, batch-upload metadata, harvest publisher metadata, and manage users.

Map Viewer

The portal map viewer shown in Figure 4 is a mapping application that allows users to view one or multiple internet map services at the same time in their web browser. Access to selected federal, state, and local Web Map Services (WMS) using the "add service" menu is provided, but this limited number of services can be expanded by entering other map server URL addresses to access other WMS available online. Viewing internet map services through the portal map viewer allows users to:

- add map services from the portal and other map servers
- display one or multiple map services in a single map view
- set the transparency of map services for overlaying multiple images
- turn map layers on or off within a map service
- find latitude/longitude anywhere in the state for automatic navigation of the map
- find street addresses in the state for automatic navigation of the map
- identify attribute information about features in a map service.

The portal is not limited to just ArcIMS WMS; it also supports several specifications and services of the Open Geospatial Consortium (OGC). The OGC is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location-based services (OGC, 2006). The portal supports the following specifications from the OGC:

- Web Mapping Services versions 1.0, 1.1, and 1.1.1
- Web Feature Services version 1.0.0
- Web Coverage Services version 1.0.0

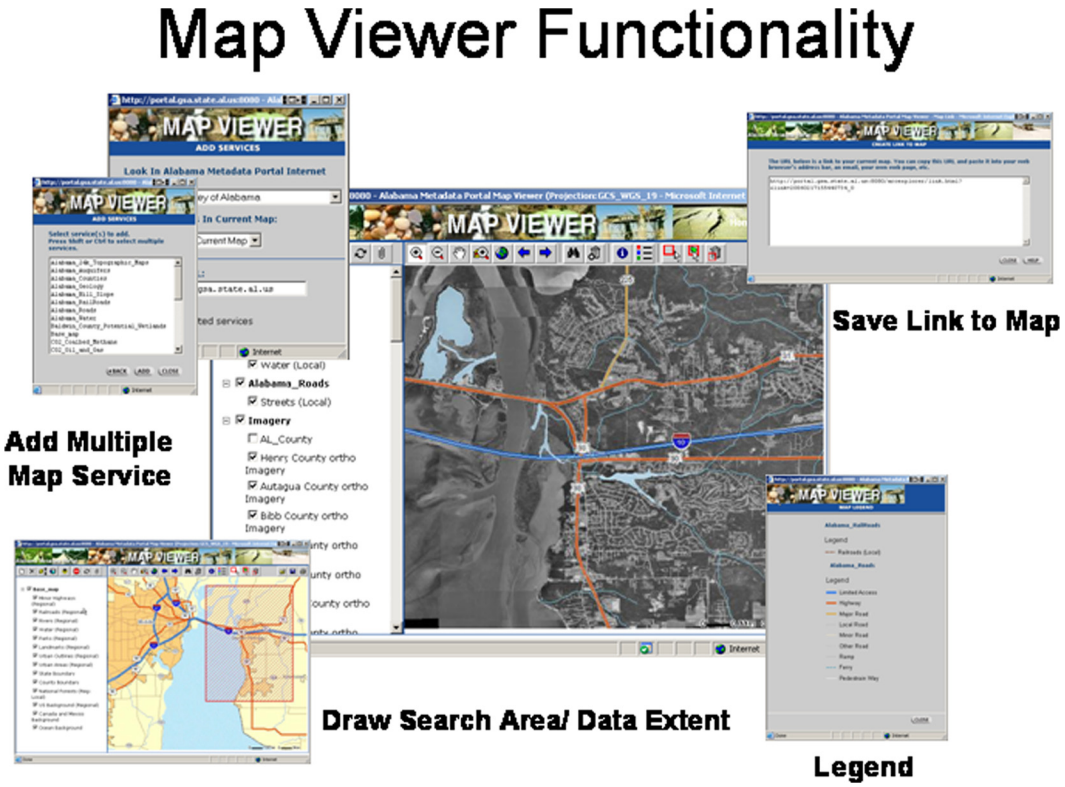

Figure 4. The portal's map viewer.

- Web Map Context Documents version 1.0.0
- Geographic Markup Language versions 2.0 and 3.0 (when approved)
- Open GIS Location Services version 1.0.

Search Function

The search page shown in Figure 5 is the tool for searching and discovering the metadata of content offered by many publishers of the Alabama Metadata Portal. The search page allows users to specify the geographic extent, keywords, content type, or content theme criteria to find matching metadata of map services, data, maps, web services, activities, or documents published in the Alabama Metadata Portal. Users can search the portal by defining "where" they would like to search, "what" in the state they would like to search, and "when" they would like the content they are searching for to have been created or updated. Users only need one parameter for a simple search; however, each additional parameter helps to narrow or retrieve a search.

THE METADATA PUBLISHING FUNCTION

The importance of writing good metadata is difficult to communicate to potential publishers of the portal. The success of the connection in Figure 1, however, is based on accurate and current metadata. Metadata describes the who, what, when, where, why, and how questions about the data, which gives users the knowledge to decide whether the data is appropriate for their desired application. Writing good metadata also mitigates the overall burdens and cost of data maintenance. The standards for including metadata records in the portal are the Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM).

There are three user levels that have metadata administration: Publisher, Channel creator, and Administrator. The administration of metadata includes the ability to create, manage, and add metadata to the portal. There are three options to make metadata records available for search in the portal. The first option is to publish a

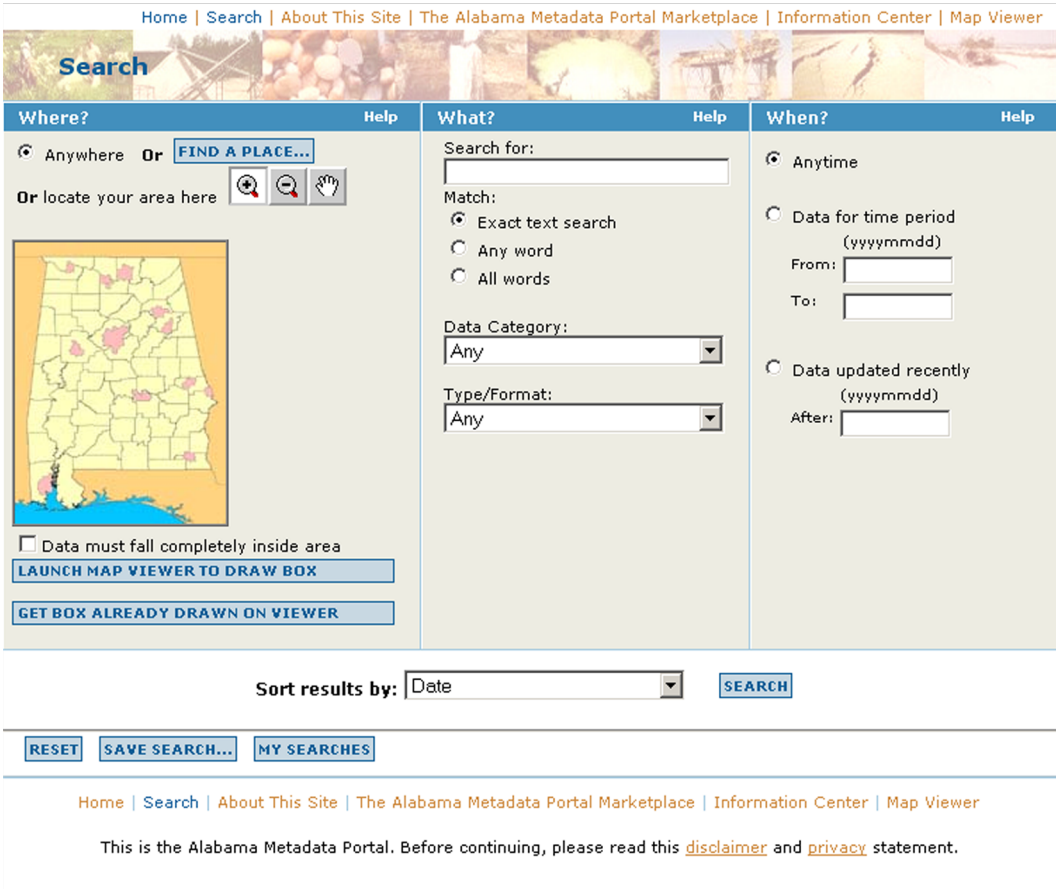

Figure 5. The portal's advanced search page.

metadata collection to a metadata repository where the portal can harvest it. The second option is to upload an individual or batch Extensible Markup Language (XML) formatted metadata record to the portal. The third option is to create a metadata record online using the portal's metadata creation tool.

Metadata Harvesting

Metadata harvesting is a self-regulated, scheduled process for collecting new and updated metadata from various metadata collection libraries. The process of harvesting allows the portal to synchronize its metadata repository with the publisher's metadata catalog. If publishers participate in metadata harvesting, any updates made to their local metadata collection will be updated in the portal during the next harvesting session. Currently, the portal can harvest FGDC-compliant metadata from three different types of harvesting protocols: Z39.50 metadata clearinghouse node, ArcIMS metadata service, and Web Accessible Folder.

Metadata harvesting in the Alabama Metadata Portal is performed in three steps as shown in Figure 6:

1. Harvesting: Based on harvesting protocol specified at the time of registration, the portal will connect to the user's local metadata repository and retrieve all new and updated metadata records.
2. Validation: During validation, the portal administrator examines each metadata record to confirm that minimum portal requirements are met. Records that are rejected are sent back via e-mail with a list of invalid fields that need to be added. The records will not be added until the metadata record is corrected and revalidated.
3. Publishing: All successfully validated and accepted metadata is published in the portal database. Once the metadata is published, it is searchable through the portal's search interface by all users.

Direct Metadata Upload

If users do not have access to any of the metadata distribution server protocols as described above, they can upload their XML-formatted metadata records directly to the portal. A metadata publisher can, through the online administration tool, add and manage metadata on their homepage. Selecting the "Upload Metadata" button, users can upload individual metadata records saved on their local computer. These records will be validated and either rejected or published in the same process as metadata harvesting. A drawback to the direct upload option is that uploaded published metadata is not linked to the local metadata repository. That is, updates to a local metadata record must be uploaded or manually changed because they are not automatically updated by the portal when the user updates local records.

ArcCatalog Direct Metadata Upload

Batch uploading of metadata records directly to the portal's metadata IMS service is possible if the user is using ESRI's ArcGIS suite. Through ArcCatalog, the user will directly connect to the portal's ArcIMS metadata server; a metadata publisher account name and password must be specified. With this connection to the portal in place, the users can drag and drop their folder of meta-

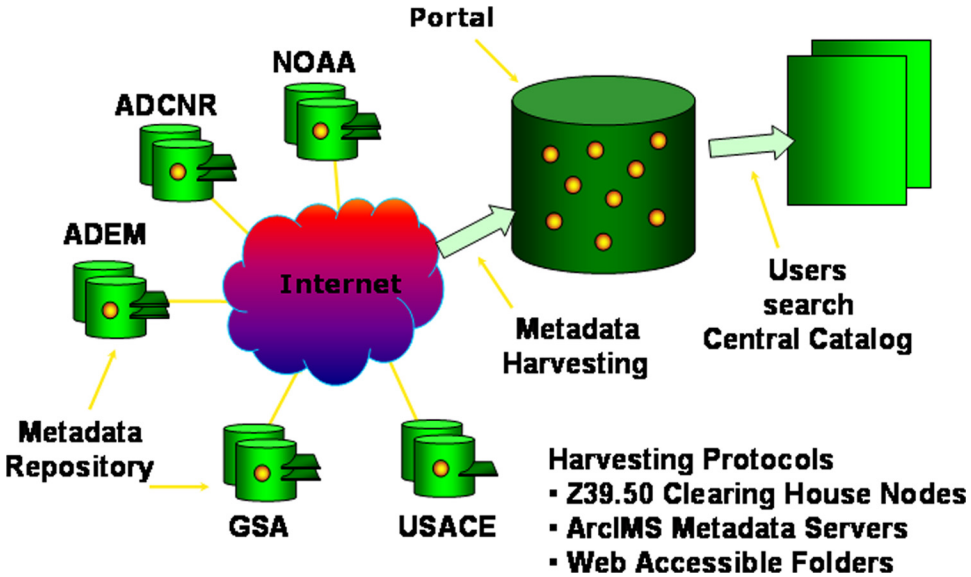

Figure 6. Diagram of the harvesting process(Modified from ESRI, 2004).

data records into the Publish Metadata Service. An added benefit to this drag-and-drop method is that the metadata record is validated automatically and displays an error message for all incorrect field values. The drawback of this method is the same as the direct upload option where uploaded published metadata is not linked to the local metadata repository as they would be with harvesting. Updates to a local metadata record must be uploaded or manually changed because they are not automatically updated by the portal when the user updates local records.

Metadata Direct Entry

The user might not have access to metadata creation or editing software, or may have very few records to contribute to the portal. If this is the case, the user can utilize the metadata creation tool provided on the home page. Users will login to their account and find the "publish online form" button under the "My Function" section. This button will take users to an online form designed to assist users in the development and production of FGDC metadata quickly and efficiently. The form provides the users with drop menus, fields that are required (indicated by *), as well as help definitions and suggestions for each of the requested metadata fields.

The minimal compliance of the direct entry method provides only the elements necessary for data discovery and is only moderately functional to users searching for data. The direct entry is a means by which to encourage users to write metadata in the hope that they will see its importance and progress toward creating a comprehensive FGDC-compliant record in the future. By using the online creation tool, the metadata will be stored only in the portal, and all updates must be made through the portal.

CONCLUSION

Data download sites and web applications have dramatically improved the GIS productivity. To complete jobs faster, it is critical that the GIS community share data effectively and efficiently: the portal is a powerful tool that benefits all users and addresses these needs. Faster discovery of specific datasets and projects, data access to download sites and use in the online Map Viewer, lowering of data costs by reducing the redundancy of data, comparison of multiple providers to find data that suits their needs, and improvement of data quality and coverage with a constant updating of agency metadata are a few benefits available through the portal. More importantly, the portal heightens the visibility of participating organizations by displaying the quality and quantity of their data offerings, which is an indication of their GIS capabilities. This allows a better understanding of how an organization could partner for future projects or initiatives.

The first 18 months since the activation of the Alabama Metadata Portal, there were 378,225 total domain hits, which represent 16,197 visits by 5,544 unique users (unique IP addresses) shown in Figure 7. We estimate that each return user has viewed an average of 68 pages. This current assessment shows the effectiveness of the Alabama Metadata Portal and the public's interest in accessing the data provided. It is important to note that the portal initiative is by no means the sole solution in producing an integrated GIS community; the portal represents a fundamental step moving Alabama into the next generation of GIS productivity.

REFERENCES

Environmental Systems Research Institute, Inc. (ESRI), 2004, GIS portal technology, an Environmental Systems Research Institute, Inc. White Paper: Redlands, California, 9 p.

Geological Survey of Alabama (GSA), 2006, Geological Survey of Alabama Portal Project: Alabama Metadata Portal, accessed 01 Sept. 2006 at <http://portal.gsa.state.al.us>.

Open Geospatial Consortium, Inc. (OGC), 2006, Welcome to the OGC website: accessed 01 Sept. 2006 at <http://www.opengeospatial.org/>.

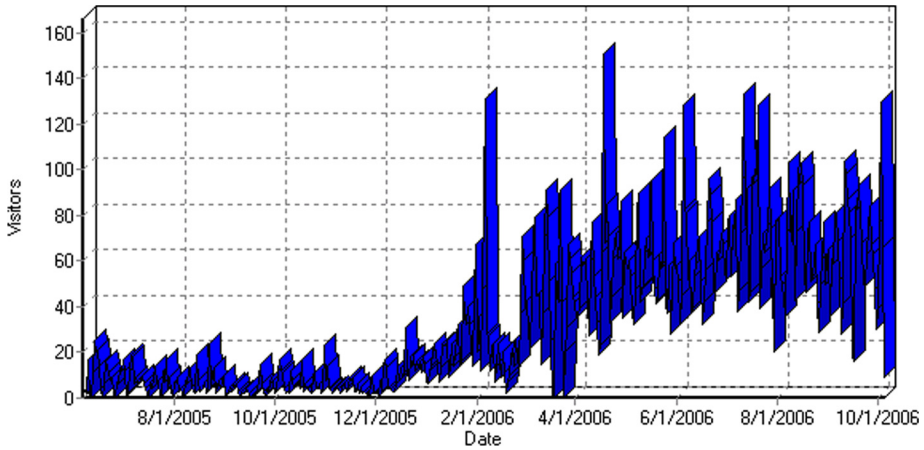

Figure 7. Daily hits on the portal from 07/01/2005 – 10/01/2006.