Development of an Interactive Web Interface for Spatial Data Warehousing

by

Nathan K. Eaton

Abstract:
Finding a specific file to download is an inherent problem with many online spatial data clearinghouses. This is particularly true for locating files that are part of a tiled data set, such as digital orthophoto quarter quadrangles (DOQQs) or local, high-resolution aerial photos. Interactive index maps used in conjunction with dynamic, database driven Web pages can provide a flexible, yet simple solution to this problem. This paper presents the methodology for developing the Indiana Spatial Data Portal (ISDP), an online spatial data distribution system, which utilizes ImageMapper (an ArcGIS extension by Alta-4) and dynamic, data-driven Web pages based on Microsoft SQL Server and Macromedia ColdFusion. The ISDP provides a single, interactive interface for downloading DOQQs, DOQ mosaics, digital raster graphics (24K, 100K, and 250K), and county mosaics from the National Agricultural Imagery Program. Discussion will include functional aspects and limitations of ImageMapper along with database design decisions and ColdFusion coding options.
The ISDP is available online at www.indiana.edu/~gisdata

Background

The idea for an online data distribution system for Indiana data was formed during the summer of 1999. Indiana was undergoing an unusually hot, dry summer, and wildfires–although small by western standards–were breaking out in the woodlots and forests of southern Indiana. State fire officials were faced with the challenge of controlling the wildfire threat and were somewhat hampered by a lack of available data for developing useful GIS maps.

About this time, the first U.S. Geological Survey (USGS) digital orthophoto quarter quadrangles (DOQQs) were arriving in the Natural Resources Conservation Service (NRCS) headquarters in Indianapolis. The Indiana DOQQ data set, which ultimately took the USGS two years to complete, is comprised of nearly 3,000 individual images. Each image was delivered in band sequential (BSQ) format and was approximately 40 MB in size. The entire data set occupied over 120 GB of disk space and required nearly 200 CD-ROMs to make the data transfer from USGS to NRCS.

Researchers at the School of Public and Environmental Affairs (SPEA) at Indiana University in Bloomington were some of the first to use the DOQQs and were soon using the files for a project to support the Indiana State Fire Coordinator's efforts to fight wildfires in the state. Experience gained from this project was the impetus for creating an online system to distribute Indiana's DOQQs.

The System Foundation – Indiana University Massive Data Storage System
Anna Radue, then GIS Lab coordinator at SPEA, forged relationships with NRCS and the University Information Technology Services (UITS) department at IU to develop the system for archiving and distributing the DOQQs. NRCS provided the original DOQQs, UITS provided

access to their newly established Massive Data Storage System (MDDS), and Radue processed the files, loaded them into the MDSS, and wrote web pages to provide a front-end for the system.

The system configuration is unique in that the data are all stored on a high performance storage system that includes a relatively small spinning disk front-end with a gigantic tape back-end. The MDSS was created to serve research data storage needs at IU and has the capacity to store thousands of terabytes (petabytes) of data. The system was also designed to provide long-term (>20 years) data storage, retrieval, and distribution. The huge capacity of the system made it possible to archive the original-resolution DOQQs as well as compressed, MrSID-format images. This is significant because as we shall see, having multiple copies of files in various formats can pose challenges for locating and downloading files of interest.

Inherent Problems with Data Clearinghouses
Data clearinghouses often provide inefficient, means of locating desired files for download. This is especially true for clearinghouses that serve data sets comprised of many hundreds or thousands of distinct, but related files. For example, a high-resolution orthophotography data set for a city or county may include several hundred individual files. In some cases, files are listed in folders that users must navigate. Other times, files are presented for download via automatic hyperlink from of an index map. Neither of these options is ideal. In the case of the folder-based example, putting files in folders requires categorizing files either alphabetically, or by using some geographic based scheme. This opens up the possibility that users will be faced with inefficiently searching folders with many files or they will be stuck with a non-intuitive folder structure that is difficult to navigate. Given the limitations of directory based file storage/navigation, in most cases it makes sense to employ index maps to help users locate files.

Limitations of the Initial Indiana Spatial Data Portal Web Interface
Over time, more and more data sets were added to the Indiana Spatial Data Portal and new users began to access data from the Portal. As the ISDP user base began to expand, more and more people began ask questions about how to locate needed files. Many people looking for DOQQs were baffled by the ISDP convention of storing the images in folders based on the USGS 1:250,000 and 1:100,000 quad names. Figure 1 illustrates the use of USGS topographic names for file management. Also, several people had trouble deciphering the cryptic names that USGS used for their digital raster graphics files (digital topographic maps). Their convention uses a

Figure 0. Example of a directory structure based on 1:250,000-scale USGS quadrangle map names.

numeric code based on geographic coordinates to differentiate files. These codes bear no resemblance to the place names used by USGS to differentiate their topographic maps. Therefore it is difficult to find maps by file name. For example, the 7.5-minute USGS topographic map for Columbus, Indiana is called "Columbus," whereas its corresponding DRG file is named "o39086b5.tif."

Redesigning the Indiana Spatial Data Portal

By 2003, with the size of the Indiana Spatial Data Portal data archive approaching 400 GB, it was clear the web interface for the Portal needed to be improved. The exiting site lacked index maps, presented files using non-intuitive geographically based folders, lacked consistent data set descriptions, and relied upon cryptic USGS file naming schemes. The site needed a major overhaul to provide the desired level of usability and functionality. The first step was to develop effective index maps for the site.

Index Map Options

There are a variety ways to create index maps for use on the web. Static images, html image maps, and dynamic map server applications are probably the most common strategies for presenting index maps. The most basic index maps, static images, provide only a visual reference to help determine a file's name and geographic location. Static images may also be useful in 'decoding' non-intuitive, geographic-based directory structures.

More sophisticated HTML image maps are simply static images that include "hot" or hyperlinked areas that can be used to link to additional images, web pages, or downloadable files. HTML image maps provide both visual and "active" file locating assistance, but they can be difficult to create and typically offer significantly less functionality than more sophisticated dynamic map server applications.

The most robust index maps are created using dynamic map server applications which generate custom index maps on the fly based on requests from web clients. Map server applications allow users to customize their index map by zooming in and out, panning, turning on and off available layers, viewing index map attributes, and linking to files for download. The downside of dynamic maps is that they require server-side map services which must be maintained by the web server administrator. This makes them somewhat complicated to set up and potentially more difficult to maintain than index maps which employ static images or html image maps.

The ImageMapper Solution

After experimenting with static index maps and downloadable published map files (PMF) for ArcReader, we discovered an extension for ArcGIS that automates the creation of HTML image maps. After a short evaluation period using a demo version, it was clear that the ImageMapper software from Alta4 (http://www.alta4.com/index_engl.htm) could provide an excellent solution to the ISDP index map dilemma. ImageMapper provides all the advantages of HTLM image maps while eliminating much of the work typically associated with creating image maps manually. It also mimics some of the functionality typically found in the more sophisticated map server based maps without the overhead of creating and managing map services on the web server. Because of the navigation and data inquiry tools provided by the ImageMapper maps, I tend to think of them as pseudo-interactive maps.

ImageMapper is a wizard-based extension for ArcMap that works by creating a series of html image maps and web pages that can be used as stand-alone web pages or easily incorporated into existing web sites. ImageMapper runs on the desktop along with ArcMap and exports a suite of JPEG images, HTML files and java script files, which can be copied to a web server to make the maps accessible on the web. ImageMapper guides the user through the process of creating a suite of maps and allows the user to specify the size of the images created, customize layer interaction (map tips and on-click events), specify map appearance (legend, scale bar, overview map), determine web page characteristics (page layout/format, attribute window size, JPEG quality), select and map navigation controls.

Once the above options have been chosen, the user specifies where to output the images and html files. When satisfied with all the options, click the Publish Map button. This generates the custom-formatted HTML files and images. The necessary files are saved to a folder which can reside on a local computer (for testing) or be moved to a web server for distribution via the Internet. Figure 2 shows an example of the ImageMapper graphical user interface.
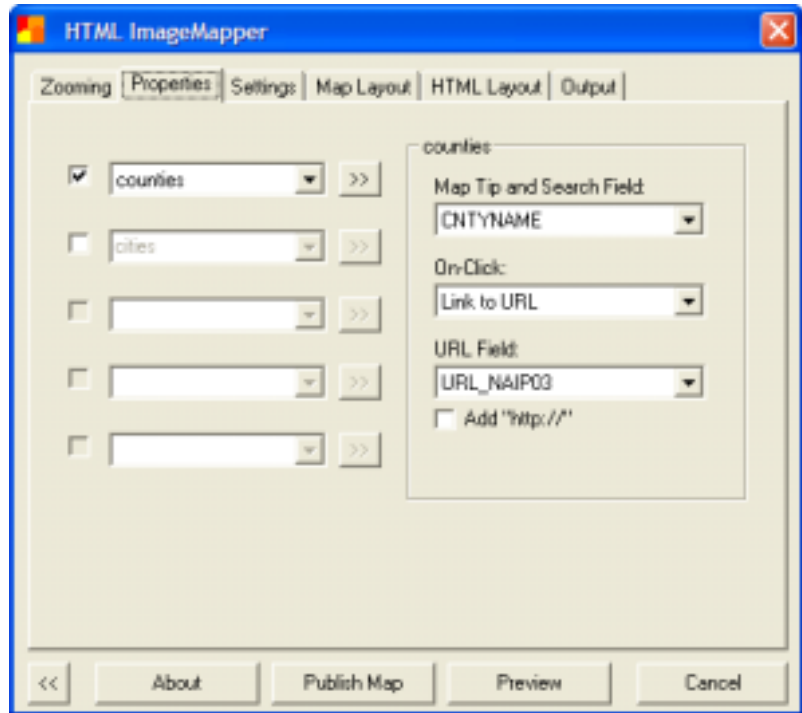

Figure 0. Wizard-based interface of HTML ImageMapper software.

Once the images and corresponding HTML pages are written, web developers can use a variety of techniques to customize the look and feel of the pages. Because we were already using Microsoft FrontPage to manage our web content, we opted to use that environment to customize our index maps. By applying a customized FrontPage Theme and Top, Left, and Bottom Border files, we were able to make the index map pages created with ImageMapper fit seamlessly with the look and feel of our website (Figures 3 and 4). We also used global search and replace tools in FrontPage and Macromedia Dreamweaver to apply page layout formatting to the suite of pages produced by ImageMapper.

ImageMapper also provides tools for adding on-click events to different map layers. For the ISDP's USGS topographic map indexes, on-click events are set to first zoom to the extent of a county. At the county scale, the on-click event changes from a navigation tool (zoom in) to a hyperlink tool which links to a new page showing all files available for download for a particular quadrangle area. The on-click events are simple to implement and work based on attributes from the GIS layers from the parent ArcMap document.

There are clear limitations in using HTML image maps for online interactive mapping, but for the ISDP application the advantages outweigh the disadvantages. One limitation of HTML image maps is that they are static images—they are not created on the fly by an application or web server. This can present a significant cost if the maps contain data that must be updated frequently. Nearly all of the data provided by the ISDP are static. In this case, the perceived disadvantage of using static images is actually an advantage. The suite of ImageMapper files is created using a desktop computer running ArcGIS and ImageMapper software. Once those files have been formatted and transferred to the web server, there is virtually no administration or maintenance required to keep the index maps functioning correctly. The map pages should continue to work as long as the web server is maintained. This may not be the case for complex interactive maps based on live map services.

Building a Dynamic Data Interface

Quality index maps are only one component of an effective data distribution system. An efficient data clearinghouse also includes thorough, standardized, data set descriptions, metadata, and intuitive file listings.

Dynamic, database-driven web pages provide perhaps the most effective way to manage the task of presenting lists of available files. By searching through database records to present only files of interest, dynamic pages can provide the end-user with a consolidated list of files, based on the results of a custom query. Database-driven pages provided a simple mechanism for cutting through complicated directory-based file management systems and non-intuitive file naming conventions that were in place in the original ISDP.

Figure 0. Sample dats set information page from the ISDP.

Although several software and server solutions are available for creating database-driven web pages, the general mechanisms by which they function are fairly similar. There are basically three components to a system utilizing database driven pages: a database with appropriate data, a web server with access to the database and software to interpret requests, and a web application that sends requests from a browser to the web server (Figure 5).

The ISDP uses a Microsoft SQL Server database back end. SQL Server provides a robust, enterprise relational database management system (RDBMS) with relatively low administrative overhead and good performance. The ISDP database includes tables which store information about available files. This information includes items such as file name, file size, file format, compression ratio, publication date, scale, and place name. Related tables store relationships between quadrangle-based files, county files, and small-scale USGS maps.
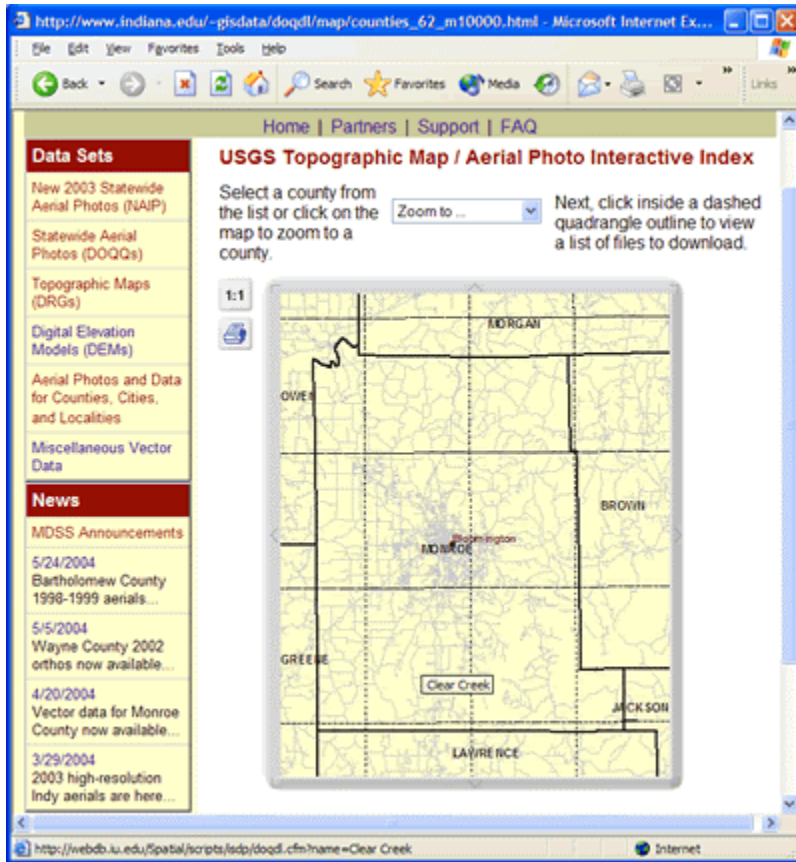

Figure 0. Sample ImageMapper index map from the ISDP.

All of the records in the ISDP database were created using Microsoft Excel and were imported into SQL Server in large batches based on data set type. Excel provides good tools for formatting data and updating large numbers of records. By using this type of wholesale data processing, the likelihood of data entry errors is reduced compared to what might be expected if inputting records one at a time. Also, considering the large number of records in the data archive (20,000), bulk data entry was the most feasible solution. we chose to use macromedia ColdFusion as the development environment for creating our database driven web pages. ColdFusion offers a custom set of HTLM tags (CFML) that allow web developers to query databases and return the results to the browser in HTML format.

Because the queries can be changed on the fly, web developers can often significantly reduce the number of web pages required for a site. For example, instead of having a separate static web page listing files related to each individual USGS quadrangle, a single dynamic page can return quadrangle-specific file lists for all 600 quadrangles in Indiana. The ISDP database driven pages present lists of available files for particular geographic areas (Figure 6). The file lists are generated from queries against a database populated with information about the files contained in the system. The pages are requested and queries are generated by clicking on index maps.
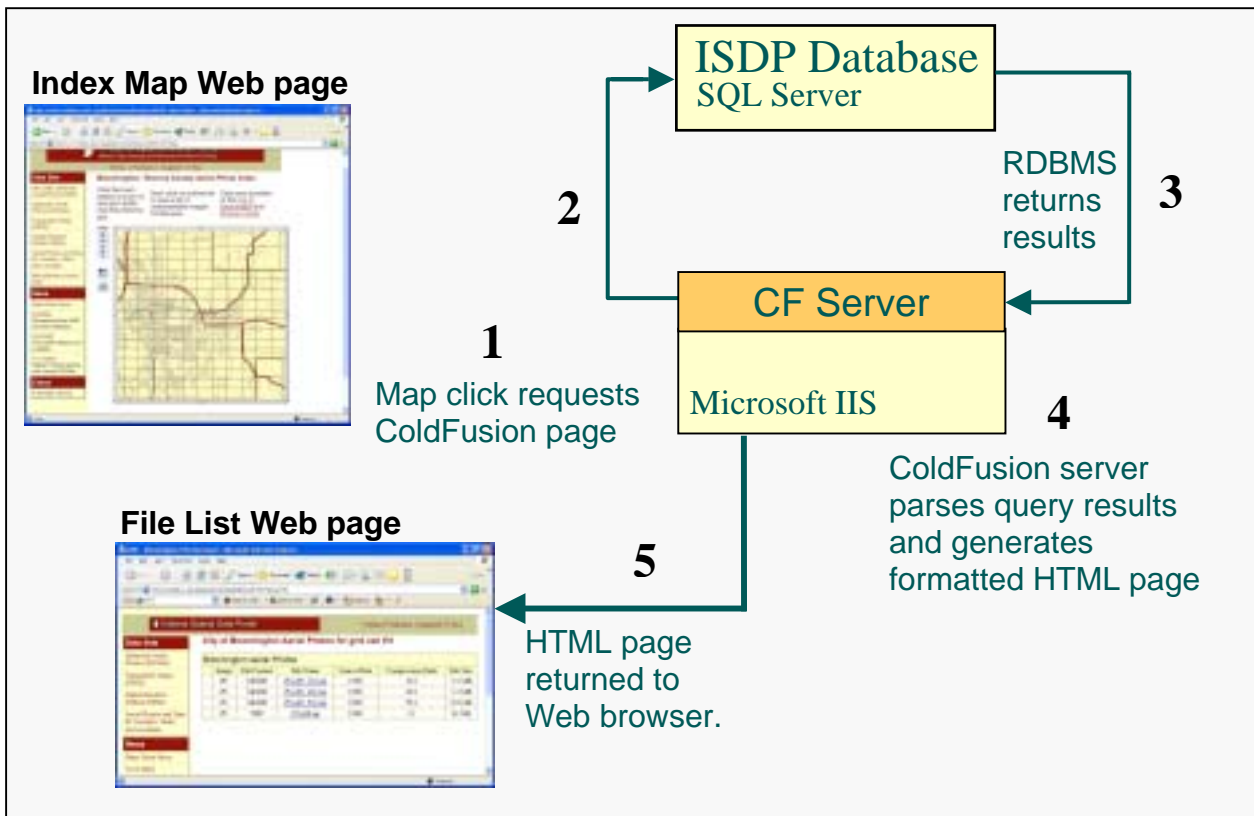
Figure 0. Diagram showing dynamic Web page process logic for ISDP.

Using the new ISDP Site

After opening the ISDP home page users select a data set of interest from the main navigation menu on the left side of the page. This opens a data set description page with provides standardized information for the data set including a basic description, a sample thumbnail image, a file format description, links to metadata, and a link to an index map for data download.

By clicking on the "Interactive index map" link from the data set description page, users can open a page containing a pseudo-dynamic index map. The ImageMapper index map provides tools for zooming in and out, panning, identifying features, and hyperlinking to other Web pages. After
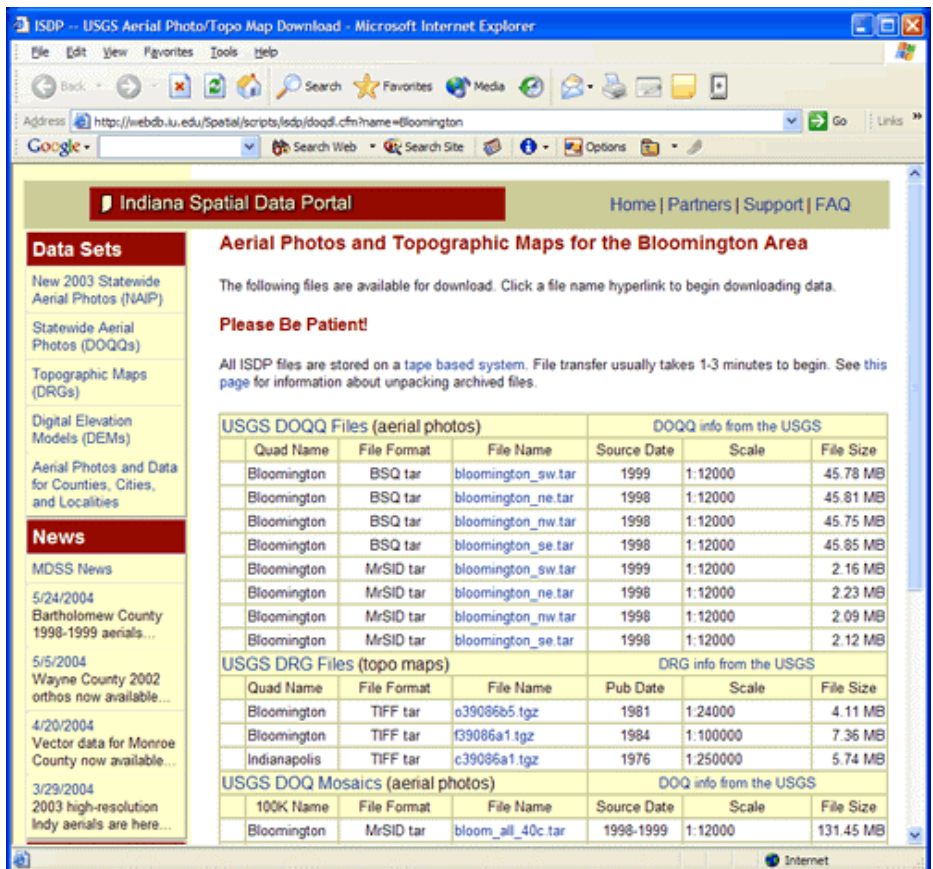

Figure 6. Sample ColdFusion page from the ISDP listing files available for the Bloomington, Indiana USGS 7.5-minute quadrangle area.

zooming in on the index map, users will see an index grid which delineates the boundary of downloadable files. Clicking on a tile of the grid opens a new window containing a customized ColdFusion page.

The ColdFusion page summarizes available files for the selected geographic area based on queries submitted to the ISDP SQL Server database. For example, clicking on a quadrangle outline on the USGS topographic map index will open a ColdFusion page showing a variety of files available for that particular area. Available data might include:

- USGS digital ortho quarter quadrangles (DOQQ) in BSQ and MrSID format,
- USGS digital raster graphic files, including 1:24,000, 1:100,000, and 1:250,000 maps,
- USGS DOQ mosaics (DOQQs mosaiced to USGS 1:100,000 quadrangle extents),
- USDA national agricultural imagery program (NAIP) aerial photos.

This consolidated listing provides download links for the geographically related files on one dynamically created web page. The user is presented with a uniform, intuitive listing of available data, even thought the files may be located in different directory structures or potentially on separate, geographically distributed servers.
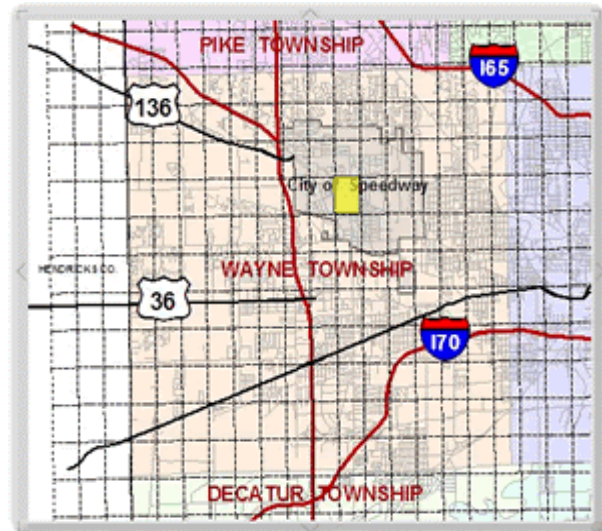
Figure 7. ImageMapper index map showing the Wayne Township portion of the Marion County, Indiana aerial photo index map on the ISDP.

The ColdFusion pages also provide the ability to show various vintages of files available for a particular area. For example, the ISDP data archive for Indianapolis and Marion County includes three years of aerial photos. By clicking on an aerial photo index tile on the Marion County index map, a user can generate a list of files that includes individual images and township or county image mosaics from 1997, 2002, and 2003 (Figures 7 and 8).

Once the ColdFusion data download page opens, users are presented with information to help them decide which files to download. The table-based layout of the ISDP data download page simplifies the process of comparing image properties and selecting the file that best meets the user's needs (Figure 8). Attributes such as source date, compression ratio, image resolution, file size, and file format are all important considerations when looking for geographic data.

Marion County Aerial Photos				data set information		
Image	File Format	File Name	Photo Date	Compression	Resolution	File Size
49_1217	TIFF (zip)	031217.zip	2003	1:1	6-inch	91.23 MB
49_1217	TIFF (tar)	021115.tar	2002	1:1	6-inch	105.07 MB
49_1217	IMAGINE	IR971115.img	1997	1:1	1-meter	4.39 MB
Marion County Aerial Photo Mosaics				data set information		
Area Name	File Format	File Name	Photo Date	Compression	Resolution	File Size
49_Marion Co.	MrSID (zip)	MarionCo03.zip	2003	400:1	6-inch	651.77 MB
49_Wayne Twp.	MrSID (zip)	Wayne03.zip	2003	40:1	6-inch	487.84 MB
49_Wayne Twp.	MrSID (tar)	Wayn2002.tar	2002	40:1	6-inch	493 MB

Figure 8. Excerpt from ColdFusion page generated by clicking on the Marion County, Indiana index map (see Figure 7). Note that available data include individual aerial photos from 1997, 2002, 2003 and aerial photo mosaics from 2002 and 2003.

Conclusions

The ISDP web site provides data storage, archival and distribution of more than 400 Gigabytes of image and base map data for Indiana. Although providing efficient access to specific files in an archive of this magnitude is not easy, standardized data set descriptions, sample images, interactive index maps and database driven file listings combine to make the process of locating any of the ISDP's 20,000 files straightforward and intuitive. The combination of ArcGIS, ImageMapper, FrontPage, ColdFusion, and SQL Server create an integrated system that is seamless and intuitive to the end-user, while being highly customizable and easily updatable for the web developer and system administrator.

Acknowledgments

Special thanks to my colleague, Anna Radue, who conceived the original ISDP and continues to be instrumental in the acquisition, processing, and archival of new data sets for the portal. Without her initiative, creativity, and drive the ISDP would simply not exist. I also acknowledge my coworkers in University Information Technology Services (UITS), Data Management Support who manage the IU Web Database Service, which provides departmental access to a web database development environment using MS SQL Server and Macromedia ColdFusion. Finally I'd like to recognize system administrators in UITS-Research and Academic Computing at IU who manage the IU Massive Data Storage System which provides the back-end data storage for the image data in the ISDP.

Nathan K. Eaton
University Information Technology Services–Data Management Support
Indiana University, Bloomington
2711 East Tenth Street, Bloomington, IN 47408
Phone: 812-855-7705 Fax: 812-856-5242
neaton@indiana.edu

Nathan K. Eaton is a GIS Support Specialist with the Data Management Support group of University Information Technology Services at Indiana University (IU). Mr. Eaton's group administers the ESRI higher-education site license for IU and maintains the Indiana Spatial Data Portal.