

Pennsylvania Spatial Data Access
The Geospatial Data Clearinghouse for the Commonwealth of Pennsylvania
Final Report
FY Ending June 30, 2004

Submitted
By
Maurie Caitlin Kelly
Penn State Institutes for the Environment
The Pennsylvania State University
August 2004

A. Introduction

The Pennsylvania Spatial Data Access system (PASDA—<http://www.pasda.psu.edu>) is Pennsylvania's official geospatial information clearinghouse and the Commonwealth's node on the National Spatial Data Infrastructure (NSDI). The PASDA clearinghouse provides for the widespread sharing of geospatial data, eliminates the creation of redundant data sets, and serves as a resource for locating data throughout the Commonwealth through its data storage, interactive mapping/webgis applications, and metadata/documentation efforts.

PASDA also serves as a primary member of the Geography Network (www.geographynetwork.com) and provides access to its metadata and data through the National Biological Information Infrastructure and Geospatial One Stop.

PASDA was developed as a service to the citizens, governments, and businesses of the Commonwealth. PASDA is a collaborative program of the Pennsylvania State University and the Pennsylvania Department of Environmental Protection. Funding is provided by the Pennsylvania Department of Environmental Protection. In 2003, additional guidance and management was provided by the Governors Office of Administration, Office for Information Technology, Bureau of Geospatial Technologies.

B. PASDA Services

PASDA data and services are provided free of charge to all users. The services available from the PASDA team include:

- **Support to PADEP**

PASDA staff have provided extensive assistance to groups and individuals within the PA DEP. This assistance includes the following:

- **Management and Technical Services.** PASDA has worked with management and technical services to provide support for:
 - GIS Town Meetings initiative throughout 1999/2000
 - SPOT Imagery access, data storage, management
 - Tutorials supporting the GIS software grant program
- **Growing Greener Program.** PASDA staff have worked extensively with GG grantees and the DEP grants center to support GIS related grants, provide data, and provide technical expertise when necessary. These grantees include the Natural Lands Trust, Pennsylvania Environmental Council, Heritage Conservancy, Clearwater Conservancy, Mifflin County, Slippery Rock Watershed Coalition/Stream Restoration Inc., Hollow Oak Land Trust, Pennsylvania Organization for Watersheds and Rivers, and others.

- **Bureau of Watershed Management.** PASDA works with the bureau of watershed management to provide documentation and storage of and access to data from the bureau.
 - **Bureau of Abandoned Mine Reclamation.** PASDA has provided extensive assistance in the development of metadata for BMR data.
 - **West Nile Virus.** PASDA is working to support the efforts of the West Nile Virus program and will be working with the WNV ArcIMS team in 2002.
 - **Office of Information Technology.** PASDA has provided extensive support to DEP OIT. Some of these include:
 - **Geospatial Data Center.** PASDA has supported GDC activities extensively both through data and metadata efforts. PASDA has provided customized data products, metadata materials, and related support.
 - **Bureau of Program Integration and Effectiveness.** PASDA has supported PA DEP efforts in the Chesapeake Bay Region by providing information to the CIMS committee.
 - **Field Operations.** PASDA has worked with the PA DEP watershed coordinators and conservation district watershed specialists providing educational materials, GIS tutorials, and seminars on GIS and PASDA to assist them in their efforts.
- **Advisory role to PAGIC**
PASDA serves in an advisory capacity to Pennsylvania state government entities in relation to their GIS and data needs under the auspices of the Pennsylvania Geospatial Information Council. PASDA serves as a bridge among agencies, projects, and non-governmental organizations to promote cooperation and coordination in the Commonwealth. PASDA staff work to further knowledge of GIS, data, and cooperative opportunities throughout the state and region.
 - **Data Liaison Activities**
PASDA staff work directly with state agencies to identify, document, and provide access to data.
 - **Inventory and Documentation (Metadata)**
PASDA will create FGDC standard metadata free of charge for any agency, organization, or data stakeholder. The metadata will also be hosted by PASDA for the purposes of developing the state data inventory.
 - **Data Storage and Access**
PASDA works directly with state and federal agencies, local and regional governments, non-profit organizations, and academic institutions to provide access to the widest variety of data possible. There is no fee to store or provide access to data via the PASDA clearinghouse.
 - **Metadata and PASDA Training**
PASDA staff provide free metadata training and training in the use of PASDA and PASDA data for individual organizations and groups. PASDA offers formal training twice a year at different locations throughout the Commonwealth. PASDA

provided metadata training to Allegheny County, DCNR, and others in the past year.

- **Educational Outreach**
PASDA works with non-profit organizations such as watershed groups and K-12 schools to promote the knowledge and use of GIS. PASDA develops lessons, tutorials, and guidelines for new GIS users.
- **Presentations & Seminars**
PASDA staff offer seminars and presentations at meetings, conferences, and schools.
- **User Assistance**
PASDA is committed to providing timely user assistance with the PASDA site and data. PASDA fields between 10 to 25 requests for information, assistance, and user inquiries per day. PASDA provides full time user assistance via e-mail and assistance via phone Monday thru Friday from 9-5.
- **PASDA Data on CD-ROM**
All data on PASDA can be obtained on CD-ROM through a cooperative arrangement with the Pennsylvania State Data Center for a nominal charge.

C. PASDA Staff

- **Maurie Caitlin Kelly**, PASDA Coordinator
- **Gary W. Petersen**, Co-Director, Office for Remote Sensing of Earth Resources (ORSER), Environmental Resources Research Institute
- **Rick L. Day**, Assistant Professor of Soil Science and Environmental Information Systems, College of Agricultural Sciences
- **Chris Pfeiffer**, Metadata Coordinator
- **Ryan Baxter**, Information Technology Coordinator
- **Tracey Walrath**, Education Coordinator
- **David Walrath**, Research Assistant & Applications Programmer
- **James Spayd**, Watershed and Conservation Resources Coordinator
- **Scott Dane**, Data Manager
- **Ryan Leech**, GIS Specialist
- **Rich Straub**, GIS Specialist

D. Major Accomplishments

1. Strategic Planning

Since the establishment of Pennsylvania Spatial Data Access as the GIS data clearinghouse for the Commonwealth of Pennsylvania in 1999, the website, services, data, and tools have expanded exponentially. The initial data and metadata offerings of

the clearinghouse included only 36 datasets and metadata records. As the result of an extensive “User Requirements and Needs Analysis” undertaken in late 2000, the PASDA team began developing unique tools and applications, and embarked on an active outreach program focused on data partnerships and education. By late 2003, these efforts had such exceptional results that the data available through PASDA expanded to 36,000 pieces of data and metadata records. In addition, a number of applications, tools, and search mechanisms had been developed to assist the ever-increasing user base (from only a few thousand a year to thousands every month) in accessing and using GIS data.

The User Centered Interface initiative (referred to as UCI) is being undertaken by the Pennsylvania Spatial Data Access team in an effort to address comments and suggestions from users, improve the functionality, navigation, and responsiveness of the website/clearinghouse, and to provide additional data search and retrieval options for users. The primary goal of this initiative is to streamline and simplify both the interactions of users with the site and data as well as the management of the data/metadata stores, website, and applications.

Background

The entire PASDA staff met for two days in late 2003, to discuss current roles and future directions for the organization, its users, and its staff. The meeting was intended as an opportunity for current staff to discuss their role within the organization; a means of anticipating how those roles might change over the next few years; and a chance to re-focus and streamline the PASDA web site and data access portals. The following outlines a set of comments, recommendations, and objectives to improve PASDA’s interface and operations:

Website

- The current web site is difficult to navigate. This is not necessarily a byproduct of the last design iteration, but rather a result of the degree to which PASDA has grown in the past three years.
- There are numerous search mechanisms, yet no topical organization or integration.
- While some of the main headings might remain the same, the web site should be streamlined and simplified to make the data search, acquisition and troubleshooting process more seamless.
- Search results and metadata information should be consistent and uniform.
- Graphics should continue to be minimal for ease of loading.

Users

- There are two major groups of people that PASDA should be targeting – ‘Expert Users’ seeking data to do analysis or make maps; and ‘Non-Expert Users’ seeking maps or other pre-formatted information.

- The Expert User group can include:
 - PA state government employees/GIS specialists
 - Federal government employees/GIS specialists
 - GIS consulting professionals
 - Academic Institution Users/Grads/Faculty
 - Utility company employees/GIS specialists
 - Local governments
- The Non-Expert User group can include:
 - PA state government employees/non-specialists
 - Non-profit conservation/watershed groups
 - K-12 educators & students
 - Undergraduate students
 - Local governments
- The site should serve both groups of visitors, but the needs of each are may not be accommodated with the same set of tools.
- Expert Users are primarily seeking data and/or metadata. There should be a set of tools that specifically targets this task.
- Non-Expert Users are primarily seeking maps and simple data search/retrieval for basic software use (ArcExplorer, Mr. Sid Viewer, etc) there should be a set of tools targeted at this task.

Data

- The types of and amount of data available through PASDA has grown exponentially.
- There is significant potential for this growth to continue over the next few years.
- The format, geographic extent, and other parameters continue to be diverse.
- The needs of data providers vary.

Applications and Tools

- Tools to access, customize data should be enhanced or developed.
- Applications should have streamlined, consistent interfaces and graphics.
- Quick access to tools and ease of use should be primary goals.

2. Data & Metadata

PAMAP Data

Data was received from the PAMAP program. The data was inventoried and the metadata was reviewed for completeness. We are awaiting permission to make this data available.

Allegheny County Metadata Training

Training for Allegheny County employees has been confirmed for October 21 & 22 at the County Emergency Operations Center.

[Pennsylvania Land Use Land Cover, 2000](#) -

Produced by Penn State University's Office for Remote Sensing of Earth Resources this layer represents a 15 class interpretation of land cover for Pennsylvania produced from LANDSAT imagery captured between 1999 and 2002.

[Western Pennsylvania Conservancy](#) -

Layers for easements, properties and locations of past projects of the Western Pennsylvania Conservancy.

[PENNDOT municipality boundary update](#) -

A problem was found with the 2004 municipality boundary file posted in February. This was a file structure problem that caused an error linking the table to geometry. This problem was remedied with the download file dated March 22, 2004.

[City of Philadelphia Zoning polygons](#) -

Addition of zoning boundary polygons to the collection of data available from the City of Philadelphia.

[American Forests](#) -

Vegetation land cover for the Delaware Valley as interpreted from 30-meter Landsat imagery for the years 1985, 1993 and 2000. Additionally vegetation land cover as classified from 4-meter IKONOS imagery captured in the year 2002 is available for the Cobbs Creek/Mill Creek and Tacony/Frankford Creek Watersheds.

[PennDOT boundary and state road centerline layers, 2004 update](#) -

The 2004 update from the Pennsylvania Dept. of Transportation. Updated layers include county, US Congressional District, PA Senate District, PA House District, PENNDOT engineering district and municipality boundaries. Also updated are the state maintained roadway layers extracted from the Roadways Management System (RMS).

[PA DEP, Bureau of Mining and Reclamation](#) -

The 2003 update to the Historical coal seam, mining, and permit data is available. Supporting database files for the permitting records have been updated. A list of quads updated since the 2001 edition is available in PDF format at:
http://www.pasda.psu.edu/data/coal/quads_updated_2003.pdf

[City of Philadelphia](#) -

Update of datasets available from the City of Philadelphia GIS Department to add vector data of topographic contours at 2-foot intervals [Note: currently the download file size is 250 MB]. Also the Philadelphia City Planning Commission census tracts for 2000 has been updated to correct an error of omission.

[City of Philadelphia](#) -

Base layer data from the City of Philadelphia. Includes street center-lines, curblines, hydrology, and other layers primarily generated at a 1:2,400 scale or better from 1-foot resolution orthoimagery captured in 1996.

Users are given the choice of downloading the native ArcINFO Coverage format or shape-file format on agreement with the terms of use.

Note: Layers in coverage format should be uncompressed using the "retain directory structure" option in your unzip utility. This will create a new directory and extract the coverage files in that directory eliminating the need to import the layer from Interchange format.

[Fractional Vegetation Cover and Impervious Surface Area layer update](#) -

Updated versions of the FVC and ISA layers for 1985 and 2000 to reflect minor changes in the estimation algorithm. Layers were estimated from data acquired around 1985 and around 2000 by the Thematic Mapper instrument on board the Landsat 5 satellite. The estimation was accomplished by processing the satellite acquired data with algorithms developed by Dr. Toby Carlson of the Department of Meteorology, Penn State University. The Pennsylvania Department of Transportation sponsored the project as part of their ongoing effort to examine the regional impacts of highway construction. The data are grids compressed in zip files which retain directory structure of the grid to eliminate the import process. The state was divided into four subareas to reduce the volume of data for downloading.

September 2003

[National Aeronautics and Space Administration](#) -

Shuttle Radar Topography Mission (SRTM) digital elevation data captured in February 2000 from the Space Shuttle Endeavor. The 30-meter horizontal resolution data is made available here in 1 x 1 degree tiles as GeoTIFF images or USGS DEM file formats.

August 2003

[DOQQ Viewer and download utility](#) -

View the most recently available USGS/PaGS Digital Orthophotos and download DOQQ's and Digital Raster Graphics (7.5 minute Topo maps) in TIFF or MrSID formats.

[Pennsylvania Game Commission](#) -

The Pennsylvania Game Commission has released updated boundary layers for State Game Land holdings and the Wildlife Management Units of PGC's Bureau of Wildlife Management.

June 2003

[Census 2000 Mapper](#) -

Summary File 3 data from Census 2000 has been added to the interactive Census Mapper. Summary File 3 (SF 3) consists of detailed tables of social, economic and housing

characteristics. This data is derived from the 'long' census form. Data is available aggregated by county, MCD, tract or block group.

[Fractional Vegetation Cover and Impervious Surface areas for 2001](#) -

Fractional vegetation cover data added to the Impervious Surface collection. Layers were estimated from data acquired between 1999 and 2002 by the Thematic Mapper instrument on board the Landsat 5 satellite. The estimation was accomplished by processing the satellite acquired data with algorithms developed by Dr. Toby Carlson of the Department of Meteorology, Penn State University. The Pennsylvania Department of Transportation sponsored the project as part of their ongoing effort to examine the regional impacts of highway construction. The data are grids in the ARCINFO exported format, compressed in zip files. The state was divided into four subareas to reduce the volume of data for downloading. Subareas 1 - 4 correspond to northwest, southwest, northeast and southeast Pennsylvania respectively.

Applications and Tools

PennCAT

PennCAT is a framework for sharing and discovering GIS data services and applications on the Internet. There is a proliferation of Internet MapService and Web GIS application development in the Commonwealth as Internet map servers such as ArcIMS become more widely used. PennCAT establishes a network that connects these data and application providers with GIS data users. PennCAT consists of a catalog of Commonwealth MapServices and applications, which can be discovered using a variety of search tools. MapServices and applications are available via a data provider's Internet map server, and users are connected directly to that source. MapServices are unique in that users can access and use them dynamically over the Internet without downloading a data file to their own computer. PennCAT is built upon ESRI's ArcIMS Metadata Server, ArcSDE, and IBM's DB2 database. PennCAT is WMS/WFS compliant.

UCI

The UCI is the new focus and core of PASDA's interface architecture. In this architecture, users will be able to add any number of data layers to a shopping cart-like collection prior to viewing or downloading them. Each other component in the new PASDA architecture, i.e., the search tools, the Web mapping tool and the download tool, communicate with the bucket to retrieve the user's data collection. The bucket will be implemented by storing a user's session information in a relational database. This information may include an auto-generated user ID, the IDs for each data layer in the bucket, a date stamp and the status of the user's session. The bucket will have a simple interface where users can view the list of data layers currently in their collection, add new data layers by returning to the search tools, or delete unwanted ones. In addition, this page provides links to the Web mapping tool, where the data layers can be viewed, and to the download tool where data can be processed (i.e., clipped and reprojected) and

downloaded. The bucket system will be implemented using ASP.NET, ADO.NET and SQL Server.

DOQQ Clipper/Reprojector

The DOQQ Clipper/Reprojector was developed to enable easy access and download for large imagery datasets. This technology will be applied to all imagery stored in the database.

Status: The clipping function is currently enabled. The reprojection function is still in testing and modifications are being made to improve performance and reliability. Expected date of final deployment of the projection function is in FY 2004/2005.

Dynamic Map Service--“Make Map”

The dynamic map service is being researched initially at the request of the DEP watershed coordinators and Growing Greener program for use by watershed and non-profit groups. Additional users/uses have arisen that make this application applicable in a number of settings including emergency management, local government planning, and conservation efforts.

WebGIS Interface Coordination/Integration

Currently, PASDA maintains a number of unique web gis based applications. These include: Census Mapper, DOQQ Clipper/Viewer, PA Atlas, PA Watershed Explorer, and others. Each application has its own function buttons with a unique look and interface. The current effort is to identify the best options for users and the most user friendly interface for all applications and apply them uniformly.

Metadata Reorganization/Hybrid Database Design

This task refers to the effort to migrate from the current file-based system in use on the PASDA site wherein the source XML files are indexed using the open-source ISITE package in use by many of the NSDI nodes for Z39.50 compatibility. The reorganization refers to a restructuring of the documentation scheme to provide a collection level documentation along with the file-level that currently exists. The search hierarchy will also need to be modified to allow users to search at their choice of collection or file level, i.e. search for DOQQ's as a collection or for a specific DOQQ within that collection.

The Hybrid Database will allow us to migrate from the ISITE file-based structure to a database driven search for the site and search services while maintaining the integrity of contributed metadata files. This involves parsing the metadata out to indexed database tables for rapid search and reporting while still allowing full access to the source XML

document as a binary in the database table. Also incorporated into this is the integration of the different paths to access that a single dataset may have, i.e. built into an application, map/feature service, interactive download or raw downloadable data.

Keyword, Browse, etc.

This activity is largely integrated into and builds upon the task above. This will conceive search strategies and operationalize the search and retrieval aspect of the metadata reorganization.

Search Results presentation

This task will integrate the presentation of search results for the many forms of data access available from the site.

MuniX - Municipal Intersections

How much of that watershed is in your town? What is the area of all State Parks are in your borough? This tool will help you find out.

GAPXplorer - Pennsylvania GAP Species Explorer

The GAPXplorer application was developed to facilitate exploration of the Pennsylvania GAP Analysis Potential Habitat datasets. These GAP datasets contain information on a total of 470 vertebrate species within the Commonwealth. Information associated with each species is linked to a geographically-based statewide grid, which contains over 118,000 1-kilometer square cells.

Geospatial One Stop

NBII

- Census Mapper now contains all SF3 data. Updated the Census Mapper tracking database by adding a SF column. The Census Mapper reporting page now indicates the variable, geography, county and SF chosen.
- PennCat, the catalog of mapping applications and services, interface development.
- Developed recreational (state park, national park etc) data finder and linked to state web sites.
- Penn Cat and all map services have enabled projection on the fly.

- The DOQ clipper was made operational. This allows for the identification and clipping of imagery based on a user defined extent.

- The Pennsylvania GAP data explorer was completed and ready for testing. The gap data explorer allows users to query, buffer, and interact with GAP data. This data is traditionally accessible only to those with advanced GIS software or knowledge. The gap data explorer makes the data accessible to users of all levels.
- Geodata.gov--This is the new map service portal for the Geospatial One-Stop initiative. Worked with geodata.gov administrators to make PASDA Geography Network/Penncat services will be available through the Geospatial One-Stop in August.
- Smart Conservation—worked with Natural Lands Trust to add updates to system architecture.
- Explore PA History—website and mapping application featured in USA today. Updated roads data in mapping application.

- Developed requirements and procedures for creating XML-based GIS web services with the .NET framework for new architecture.
- Made final modifications enhancements to beta version of the DOQ clipper.
- Installed and configured DOQ clipper on production servers.
- Wrote a VB class that administers ArcIMS services programmatically.
- Developed an application logging service with XML, VB.NET and ASP.NET.
- Began developing data model for the new PASDA architecture, specifically how the layer collection system (bucket) will be stored within the DB2 database.
- Prepared PennCAT for deployment. Edited the code behind PennCAT so that when a data search is performed, each dataset's originator will appear in the search results list; this did not happen by default.
- Identified first partners for PennCat deployment. Delaware River Basin Commission, PADCNr, and PADEP most suitable.
- Implemented geocoding tool in the DOQQ Viewer so users can zoom to a specific street address and zip code. This tool uses the GDT roads data in SDE/DB2. Implemented a layers on/off tool that turns all layers on or all layers off.

Outreach & Assistance

- Worked with PEMA to provide latest data/assistance with DOQQs in ArcIMS.
- Worked with PHMC/WITF to provide updated data to Explore PA History site.
- Worked with EPA project on Environmental Mark Up Language.
- PASDA presentation at the Western PA GIS Conference by the California University of PA.
- PASDA presentation at the Chesapeake Bay Program IT Meeting in Annapolis, MD
- SEDA-COG GIS Task Force Meeting in Lewisburg, PA
- PASDA exhibit at the Pennsylvania State Data Center Conference in Harrisburg, PA
- Meeting in Harrisburg with Art Stevens, Phil Coombs, Jay Parrish, Nancie Imler, Ebby Abraham.

- Attended and exhibited at the Pennsylvania State Data Center Conference

- Attended and provided update for PAGIC meeting
- Met with Nancie Imler and Ebby Abraham PADEP, Jay Parrish, DCNR, Phil Coombs, Art Stephens, and Kristen of Office of Administration

Meeting with PA DEP Water Conservation Bureau regarding watershed groups and watershed notebooks website.

Attended and presented at Western Pennsylvania GIS Conference

Attended and presented at Chesapeake Bay Information Management Subcommittee meeting.

Meeting with Slippery Rock Watershed Coalition. Attended and exhibited at annual riverfest gathering in Pittsburgh with SRWC and other watershed groups.

Meeting with PA State Data Center, Brady Stroh.

Meeting with PA DEP Watershed Coordinators, Growing Greener group to discuss PASDA's continued assistance with their efforts.

Presented at Tyrone Area Middle School.

Attended and exhibited at Clearwater Conservancy Spring Creek Day.

Meeting with Lancaster Community College.

E. PASDA Data, Metadata, & Infrastructure

FTP site

The base of the PASDA site has always been its data archive, currently a flat file structure available via FTP. A mix of raster formats including TIF, BIL, BSQ, and GRID and shape-files and coverages (in e00 exchange format). All files on the FTP site are zipped for a more efficient download. The FTP site itself is arranged by 'collection' in some cases these are arranged by provider or in other cases by the theme topic (DOQQ, floodplains, etc.).

For distribution the layers are zipped by PASDA staff with no modification from the original source. In some cases the data may be made available in a variety of formats or geographic extents. For example the watershed, floodplain, streams and roads originally produced as ArcInfo coverages in the Albers projection were transformed by PASDA and are available as shape-files in UTM and as unprojected layers to simplify use in ArcView and with the USGS imagery. Many statewide layers have been clipped and are distributed in county files to simplify download. In all cases the original data is archived and available from the FTP site.

For more current additions to the PASDA collection metadata for data layers are included in the compressed archive in both text and XML formats for distribution with the data. Text files follow FGDC content standards and XML files are validated against the FGDC Document Type.

The FTP site is available via anonymous login and is accessible through either a web browser or simple FTP client for bulk downloads. The FTP service is configured to require a reverse DNS lookup to access the site for security reasons.

Web site

The web server is a Sun 350 running Apache web server software. Data access pages are static pages created to list available data by collection allowing access to metadata, preview images and direct download of the data. Direct access to the FTP site also available through the website.

The open-source Isite web-site search suite is used for keyword search access to the data collection. Made available by the Center for Networked Information Discovery & Retrieval (CNIDR) Isite is fully supported by the FGDC for use in the NSDI. This is used to provide search access to both the metadata collection and the website pages. Metadata on the site is stored as discrete XML files making use of the XLINK specification from the W3 Consortium for redundant information. The Isite suite includes an indexing program which harvests information from the XML files for retrieval through the website. A recent upgrade of the software will allow the indexing of metadata files from remote sites to allow a search of distributed sites while maintaining the indexes locally for high performance.

Geographic search capability is made possible through three scripted image map interfaces for watershed boundaries, county boundaries, and USGS quadrangle boundaries. This search uses bounding boxes to retrieve metadata for all indexed layers that are entirely or partially included in the respective features. This capability is built into the Isite suite.

Once accessed, the XML metadata files are parsed on the fly through stylesheets for browser display, printing or saving. Links for data download and preview if available are provided in the metadata.

General Infrastructure

The PASDA infrastructure and search capabilities are based on the Z39.50 standard protocol and use Isite software in accordance with the NSDI standard. All PASDA metadata is generated in XML format using the Federal Geographic Data Committee Content Standard for Geospatial Metadata (in the interest of creating a comprehensive data inventory for the state, PASDA staff provide free metadata/documentation services for any agency or interested GIS data provider in the Commonwealth on request). PASDA data resides in two forms. First, PASDA maintains an FTP site where all data sets are zip/compressed (DOQQs are compressed first with Mr. SID compression software then zipped for ease of downloading) and accessible to anyone with an Internet connection. These data may be accessed through several types of searches (keyword, geographic [by county, quadrangle, watershed], topic, or provider) or by going directly to the FTP site. PASDA data also resides in a relational database (DB2 with Arc SDE) and is accessed through an ArcIMS Web GIS interface. This interface allows users to customize their data requests by identifying, viewing, selecting, reprojecting, and clipping desired data.

The basis of the PASDA data/metadata repository is state-wide framework data which includes digital ortho photographs, digital elevation models, digital raster graphics (scanned, geo-referenced images of the 7.5' USGS topographic maps), the state road network/transportation data, hydrography/state-wide streams, floodplains, river basins, and governmental units. In addition to this data, PASDA acquires cadastral data such as property and parcel data, digital aerial photos, and land use data from regional and local governments. Biodiversity information is included in the PASDA data repository through agreements with state agencies, such as the PA Game Commission, with non-profit institutions such as the Academy of Natural Sciences which provides the Pennsylvania Breeding Bird Atlas data, and with academic institutions engaged in Federally funded state-based research projects, such as the GAP Analysis Project.

Data available through the clearinghouse are, in most cases, stored in a central location. However, in some cases, in particular those of local government data providers, some data may be distributed, with links provided from PASDA via the metadata.

Hardware/Software Specifications

Aside from the legacy web server, PASDA maintains a development server cluster in the Land & Water Building and a cluster at the CAC. PSU provides ArcSDE and DB2 for database development purposes. Other details of the cluster machines are as follows:

ERRI Cluster

Dell 2400 Server (two 500 Mhz Processors, 1GB RAM)

Running Windows NT 4 and IIS 4

Unify ServletExec Java servlet engine

ColdFusion 4.0

ArcIMS 3

3 Dell Precision workstations

1 running AutoDesk Map Guide

2 ArcIMS spatial servers (DB2 Spatial Extender will be installed on one of these)

CAC Cluster

1 Sun Enterprise 350 dual processor server with 2GB RAM

Solaris with Apache and Jserv Servlet engine

PERL

Python

ArcIMS application server

3 IBM Netfinity dual processor servers (eventually with 1GB RAM each)

Windows NT4

ESRI ArcIMS 3 Spatial Server

Shared storage array initially with 160GB disk space

Access to CAC's IBM dual processor server running AIX and ERRI Cluster DB2

UNIX, AIX, and DB2 administration from CAC personnel