Final Report
TASK #57
Pennsylvania Spatial Data Access

Submitted by

Todd Bacastow, Maurie Caitlin Kelly
Christopher Pfeiffer, Marty Gutowski
Ryan Baxter, Tracey Walrath
Scott Dane, David Howard
Ralph Heidl

Environmental Resources Research Institute
The Pennsylvania State University
University Park, PA 16802
January 15, 2000

A. Introduction

Pennsylvania Spatial Data Access, PASDA, was named the official state geospatial data clearinghouse for the Commonwealth of Pennsylvania in September 1999. PASDA, which has been in operation since 1996 through a contract with PADEP, is also a National Spatial Data Infrastructure (NSDI) Clearinghouse node. PASDA provides free public access to geospatial data and metadata via the World Wide Web. PASDA is a collaborative effort of the Pennsylvania Department of Environmental Protection, the Pennsylvania Geospatial Information Council, and the Pennsylvania State University. It 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 metadata/documentation efforts.

The overall management of PASDA as well as the metadata, training, interactive mapping, and education and outreach efforts are the responsibility of Penn State's Environmental Resources Research Institute. The Deasy GeoGraphics
Laboratory , a unit of Penn State's Department of Geography, is responsible for web interface design, data storage, and development of data search and delivery capabilities.

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

- Metadata and data documentation services

PASDA metadata specialists will identify, document, and create Federal Geographic Data Committee (FGDC) standard metadata at the request of data stakeholders throughout the state. In addition the PASDA metadata specialists are available to assist data managers with the creation of FGDC standard metadata.

- Training

The PASDA team provides training in the use of PASDA data and in metadata creation.

- Data Storage

PASDA is committed to providing access to a wide variety of data related to Pennsylvania.

- User Assistance

The PASDA team provides extensive user assistance via e-mail and phone and in person. PASDA receives an average of 15-20 questions daily on topics as broad as watershed conservation, data availability, use with the PASDA site, metadata development, and emerging technologies.

- Online Educational Resources

The PASDA team has developed extensive FAQs and online tutorials to assist users in accessing, downloading and using PASDA data, as well as understanding issues related to GIS data.

- Outreach

The PASDA team participates in and presents at meetings, seminars, and conferences to publicize the availability and services of the clearinghouse. The goal of the PASDA team is to serve the broadest possible audience of users to maximize the cost savings related to data development and use. In addition, PASDA serves as a national resource of expertise for groups such as the Federal Geographic Data Committee and other clearinghouses by sharing knowledge, techniques, and experience in the areas of metadata and clearinghouse development.

- Consulting to the PAGIC

The PASDA team serves as consultants to the PAGIC on issues related to data dissemination, integration, standards, and access. PASDA provides extensive expertise in GIS and information storage and retrieval as well as new developments in technology.

- PASDA Data on CD-ROM

The PA State Data Center will copy PASDA data to a CD for a nominal fee.

B. PASDA Needs Analysis

Pennsylvania Spatial Data Access, PASDA, was named the official geospatial data clearinghouse for the Commonwealth of Pennsylvania on September 10, 1999. This designation was given by the Pennsylvania Geospatial Information Council (PAGIC). In light of this designation, the project staff embarked on a series of new initiatives which included the development of a broad-based needs analysis to define the services and functions the clearinghouse will offer. This needs analysis, which will be completed in May 2000, commenced in July 1999 with a series of planning meetings. Throughout the course of these meetings, two independent consultants, Advanced Technology Systems and Soza and Company, were brought in to research and implement the needs analysis. During the period from July through December, ATS and Soza developed an extensive questionnaire and conducted research into various clearinghouse scenarios.

The following information is the result of this initial research:

PASDA plays and can continue to play a vital role in the Commonwealth by:

1. Facilitating the sharing of data in the Commonwealth
2. Facilitating and assisting the state agencies in their efforts to develop a broader understanding of how their data can be used and useful. I believe these efforts in turn may help them to help themselves learn more about their own agencies and how to fulfill their data related mission.
3. Educating the citizens, businesses, educators, and organizations of the Commonwealth about spatial data, GIS,etc. And through that effort, make them more effective at what they do.
4. Assist local and regional governments in managing their own data resources (example--the metadata training sessions).

C. PASDA Goals

Concurrently with the needs analysis, PASDA staff developed and in some cases implemented a series of goals for the project. The following is an overview of these goals.

Pennsylvania Spatial Data Access is comprised of three main components: the technology that drives the system and serves as its framework; the outreach efforts to acquire data and

metadata; and the education and service component that strives to provide educational resources for non-expert users and to educate the public about the site.

Overall goals:

PASDA will become a vital and integral service to the Commonwealth of Pennsylvania. PASDA will adapt new and useful technologies and services to continuously improve its service to PA. PASDA team members will be highly responsive, responsible, flexible, and innovative in their approach to the project. PASDA will be a prized service to the Commonwealth and will be recognized for its value in terms of dollar savings, personal interactions, openess, and vitality. The PAGIC will be able to refer to PASDA as an example of the success of their efforts.

1. Responsiveness

This goal is one that relates to all others and affects all PASDA staff members. For PASDA to be successful, PASDA team members must be immediately responsive to requests for information as well as requests for action. Deadlines and benchmarks must be set and met. PASDA team members should be responsive to the requests of users, funders, PAGIC, and other team members.

2. Responsibility

Each team member has a unique responsibility as well as shared responsibilities. Everyone should be able to define and set goals for their area of responsibility in consultation with other project members and the coordinator. PASDA team members have a definite responsibility to the PADEP and PAGIC and its member agencies.

3. Cooperation/Communication

PASDA team members should cooperate and share ideas, goals, and problems with others on the project. PASDA team members should actively cooperate with citizens, organizations, governments, and businesses of the state.

4. Flexibility

Since we are in a rapidly changing environment, PASDA team members should be flexible and able to adapt to new and urgent needs and opportunities. The PASDA site should be able to incorporate new data, ideas, and functions easily.

5. Innovation

In addition to reacting and being flexible, PASDA and its team members should keep abreast of emerging trends and technologies. Information about these emerging trends and technologies will be shared, and where possible, practical, and desired, will be implemented to the fullest extent funding allows.

6. Feedback

PASDA team members are expected to provide and receive feedback from users and other team members. Our responsiveness to this feedback is vital to the success of the project.

Technology Goal: State agencies collect and automate data with different scales, accuracies, and basemap references. Not having a standardized data set of GIS data makes it difficult for state agency users to easily use the data and nearly impossible for the general public. The PASDA Clearinghouse should evolve from an online browser-based catalog and library to a system that allows both expert and non-technical users to search for specific types of GIS data and to rapidly integrate this data with their spatial data systems via the Internet. The PASDA Clearinghouse will give users alternatives tailored for their organization that preserve their investments in data, infrastructure, and expertise. To achieve this, PASDA will be reorganized around an open DBMS that provides an integrated data environment to serve both the non-technical users and developers of advanced Web-based systems.

Outreach Goal: PASDA will serve the citizens of the Commonwealth through training and outreach efforts. These will include presentations, training sessions in relation to metadata, use of the site, and use of the data, and interaction with potential geospatial data stakeholders including but not limited to state government agencies, local governments, non-profit organizations, and educators. The PASDA team will serve as a vital resource to the Commonwealth through their expertise.

Education/Service Goal: PASDA should evolve to encompass a "clearinghouse" for educational resources related to GIS and spatial data for K-12 educators and non-expert GIS users across the Commonwealth. PASDA will provide both the data and the online educational tools (or link to available tools) to enable these users to access, use, and understand GIS and spatial data. The PASDA team will be a recognized presence and resource in the Commonwealth.

D. PASDA Operations

The PASDA staff also continued to conduct its regular services during this time. In addition, PASDA participated in a variety of projects including the Pennsylvania Interactive Watershed Project and the Pennsylvania Framework Data Project both sponsored by the Federal Geographic Data Committee. The Pennsylvania Interactive Watershed project which featured a data collection and documentation initiative, a data coordination meeting, and an interactive mapping
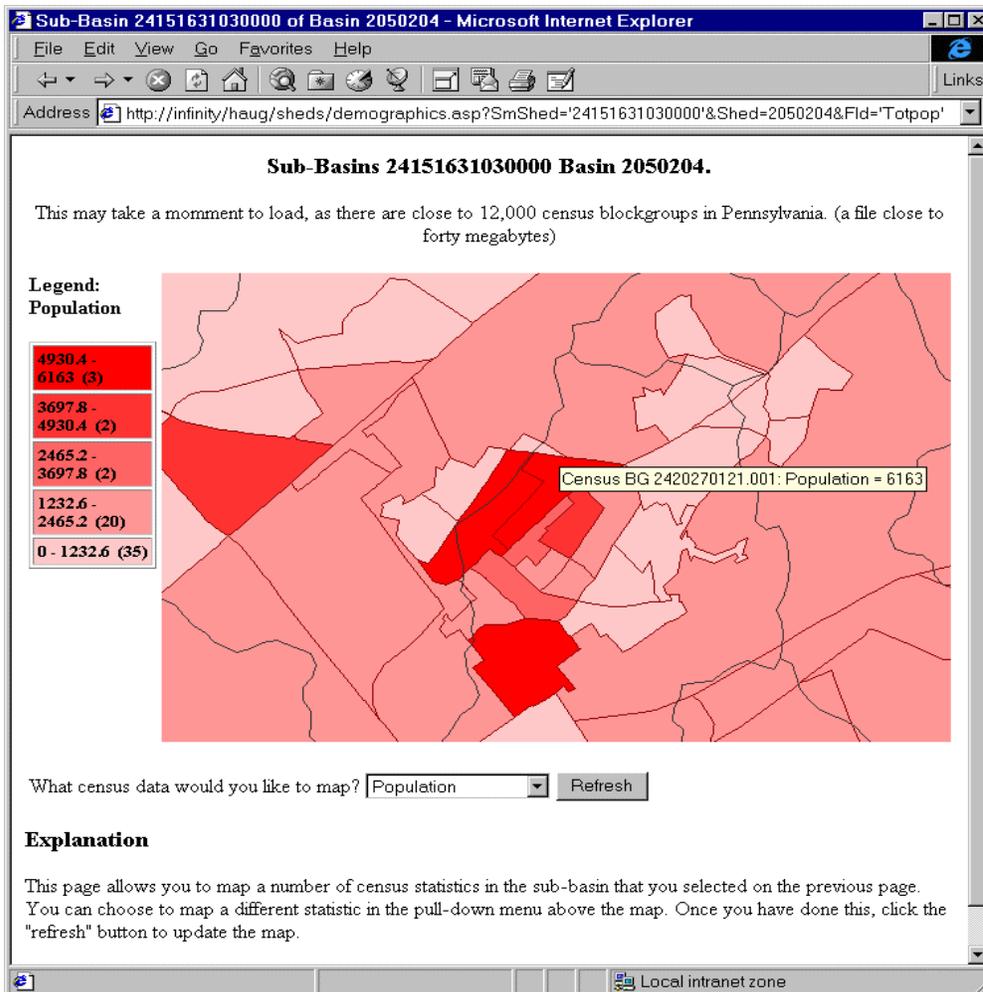
PA Interactive Watershed Atlas Prototype.

application that provided data aggregation of census block level data and stream order identification capability was highly successful. The interactive mapping application unfortunately remained unstable due to the software, GeoMedia WebMap. During the coordination meeting, PASDA was identified by participants as the most likely source for storing and distributing watershed based data.

In addition to this project, Chris Pfeiffer, PASDA metadata coordinator, developed a number of metadata related initiatives. A metadata training session was held in conjunction with the Southwestern Pennsylvania Commission in December 1999. This session focused on the new PASDA metadata form. This form was created as a "plain english" form to assist potential metadata creators. By creating the form in "plain english", it will be possible for those without metadata training to create FGDC standard metadata.

PASDA "plain english" metadata entry form.

Another initiative in which PASDA was involved was the Erie Online Geographic Information System. PASDA staff assisted in the development of this site, which was created by Jason Shenk (PSU/PADEP), as well as data acquisition and documentation. The Erie site, which is currently linked under the PA Explorer section of PASDA, was created using ESRI's Map Objects Internet Map Server.
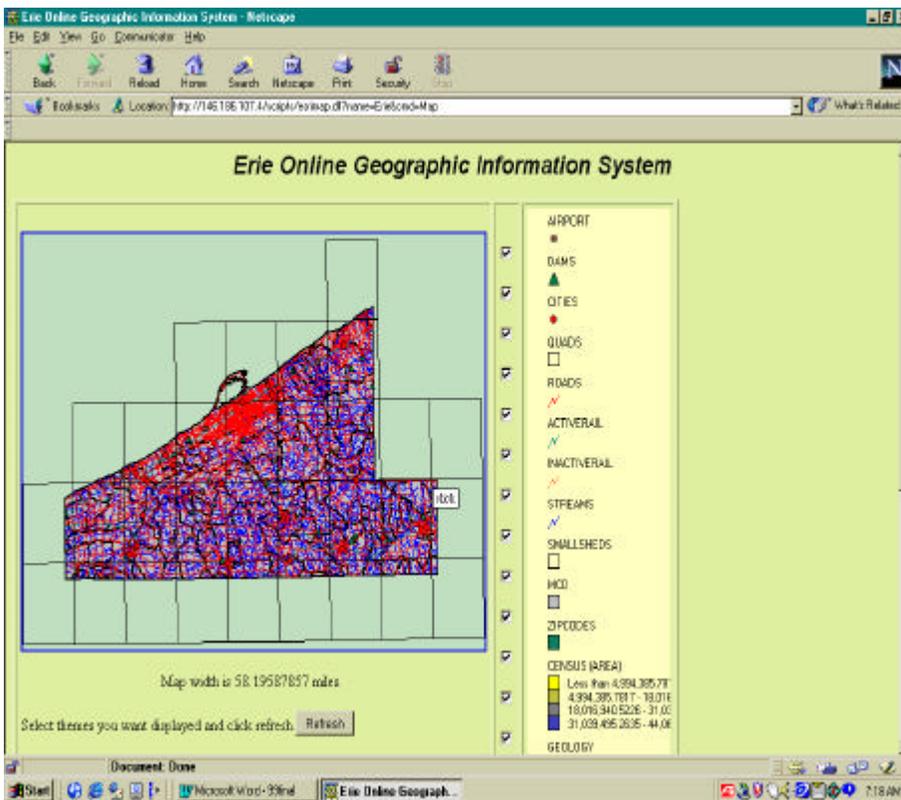
Erie Online Geographic Information System.

The Map Gallery on PASDA was another addition to the site. This "gallery" includes maps that show the physiographic provinces, recreational areas, and topography of the Commonwealth.
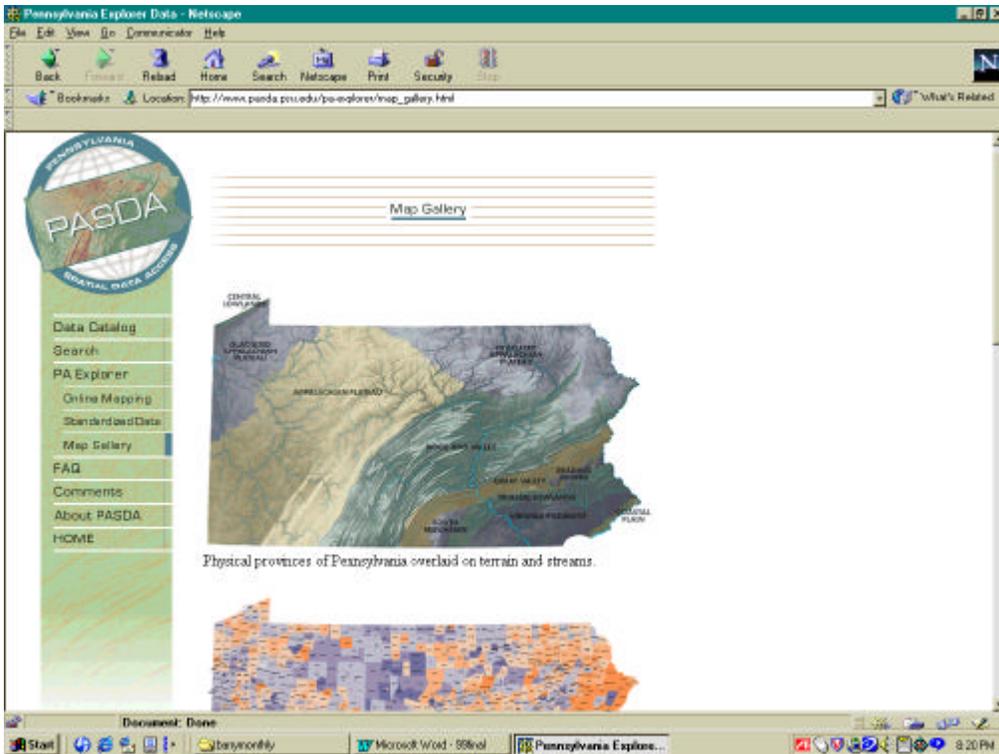
PASDA Map Gallery.

The PASDA team also developed a set of standardized data, one set in decimal degrees, the other in UTM, to accommodate our users who do not have the ability to reproject data.

PASDA Standard Data.

A significant effort to create an extensive FAQ for users was undertaken during the past six months. This effort resulted in information on projections, datums, scales, as well as explanations for common questions such as "why doesn't my data overlay (see below).
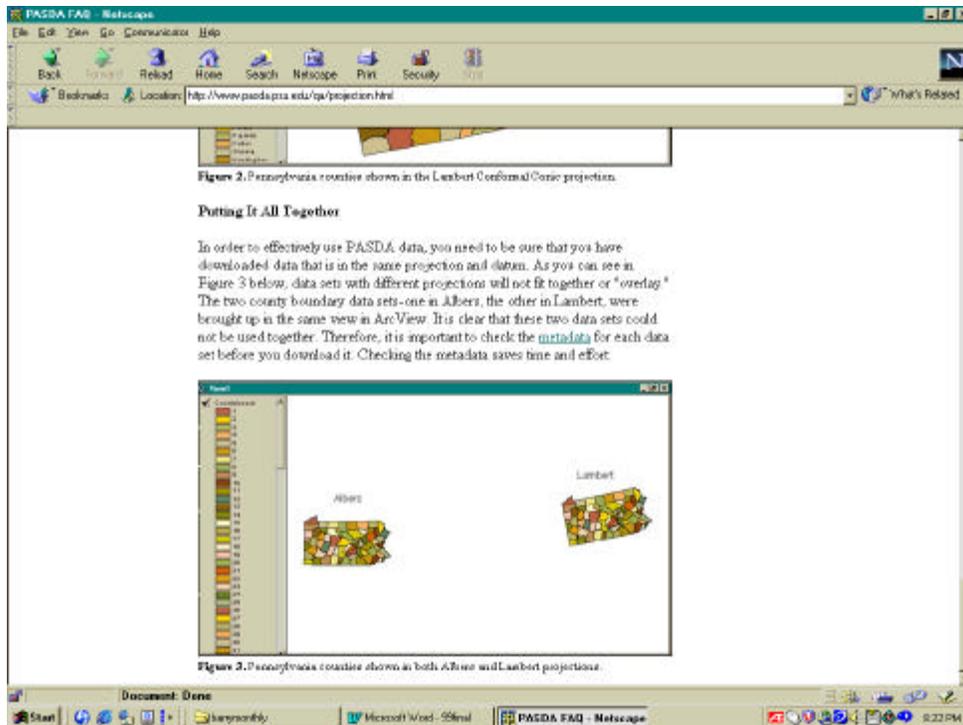
Another significant accomplishment during this time was the completion of the joint PASDA/PAGS/PennDOT/PADEP DOQQ project. This year long project resulted in over 3000 DOQQs covering the Commonwealth being made available for download and viewing via PASDA. Exceptional assistance was provided by Thomas Whitfield of PAGS in this effort. In addition, Marty Gutowski, the PASDA webmaster who spearheaded this effort, worked with Ben Lewis of PennDOT to acquire consistent GeoTiff versions of the original BIL files.
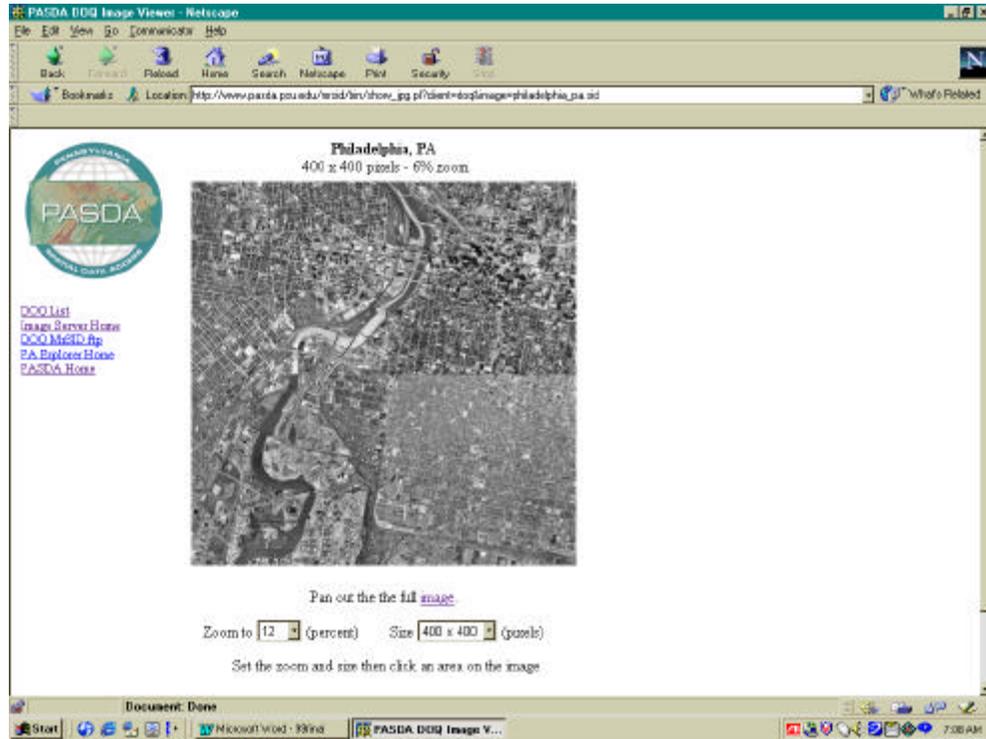
PASDA DOQs

E. Technical & Data Enhancements

- Clipped UTM DRGs, placed on ftp site.
- Created needs survey.
- Currently working with designer to create new site design.
- Continuous monitoring relevant technological advances in internet-based services
- Analyzed the leading internet-mapping products and implemented several services for product evaluation and comparison
- Designed, implemented, and tested a prototype of an internet-based orientation aide based integrating spatial data (PASDA "PSU Digital Guide" Internet Map Server)
- Reviewed autoregressive integrated moving average (ARIMA) models for forecasting PASDA customer behavior from web-activity log for July and September 1999.
- Identified areas of technology convergence related to Internet Mapping technology

- Addressed all components system administration and maintenance for the PASDA Internet Mapping Server
- Installed and configured Autodesk MapGuide 4
- Installed and configured ESRI ArcIMS beta 3
- Installed and configured secondary software (e.g. Allaire Cold Fusion 4, Jrun Servlet Engine)
- Created PA Explorer shapefile data directory in DD and UTM (403mb)
- Added 700 Digital Ortho Photo Quadrangles
- Assisted five users in fixing name server problems they had in their own organizations (the PASDA server will not accept users who do not have properly configured DNS).
- Best Practices research on search functions
- Prepared for new look and feel with rollovers and white background.
- ArcView tutorials in preparation for creating educational modules
- Edited quads coverage to correct problems in Erie County.
- Created PADSA data inventory file within excel.
- Correcting Beaver County roads from PA Explorer.
- Created a new Beaver County roads shapefile for PA Explorer.
- PASDA data inventory - filling out Excel file framework of data files.
- PASDA data inventory - filling out framework of data files.
- PASDA data inventory - filling out framework of data files and associated information.
- PASDA data inventory - adding projection information.
- PASDA data inventory - adding projection information.
- PASDA data inventory - classifying datafile types and projections.
- PASDA data inventory - classifying file types and projections.
- PASDA data inventory - classifying file types and projections.
- Opened and viewed every DOQQ to verify and fix the names
- Worked on new maps for country/quad search
- ALLARM data was upgraded
- Created county dems from the National Elevation Database (NED)files
- DRBC data added
- Version 2.0 of the MrSID server was successfully loaded, tested
- Created county and quadrangle map searches
- Created rollover watershed search page

F. Metadata

- All metadata operations were moved to the PASDA office, 141 Land and Water Building in University Park.
- Received metadata submission through the online Data Interview form to be added to the PASDA website. Based on information submitted from the Municipality of Murrysville and feedback from Chris Rearick the form was revised and is ready for release. The Form will be linked from the PASDA homepage along with other pages that will present metadata resources in conjunction with the Metadata Training Session in Pittsburgh on December 9.
- Created PASDA Data Interview form which will allow the user to submit information through an intuitive question and answer format.
  This will lessen the need to provide in-depth training, reduce the resources required to provide complete data documentation for agencies in the Commonwealth and provide a continuing education resource to state and local agencies to familiarize them with the content and structure of metadata while they document datasets. Further hands-on sessions using locally installed applications will then be scheduled for parties identified as desiring to maintain metadata information locally.

- Supporting tables for producing quad and county tiled metadata have been revised and will be extended to include information about the DOQQ's. Discrepencies between USGS information and PASDA information are in the final stages of correction. DRG and DEM metadata has been updated. Metadata for DOQQ's will be updated to include image acquisition dates and metadata for new clipped-collar DRG's will be produced using these corrected tables before the holidays.
- Met with Bill Toothill and James Thomas at the Wilkes University Center for Geographic Information Sciences. They will be providing PASDA with metadata and data from the American Heritage River Project, Tunkhannock River Conservation Plan, Luzerne County and the City of Wilkes-Barre. Initial metadata from the Tunkhannock RCP is expected before the holidays.
- Provided information on XML implementation to Roger Longhorn of the GI 2000 initiative for the European Commission and John Evans of MIT Dept. of Urban Studies and Planning. Both are at the beginning stages of implementing of the use of the XML structured markup for metadata which is used on the PASDA site.
- Contacts with Robert Cheetham of the Mayor's Office for Information Technology in Philadelphia. They are beginning to develop an enterprise-wide metadata system for all of the city agencies and were looking for information. They are interested in training on the metadata standard. Our discussions will continue.
- Contact with George Calaba of Three Rivers Connect in Pittsburgh. He is looking for information regarding metadata and the FGDC standard and organization. He will be directed to this information which will posted to the PASDA. FAQ.
- Parcels for Municipality of Murrysville
- Metadata for USGS Digital Raster Graphics reviewed and updated.
- Metadata for USGS Digital Elevation Models reviewed and updated
- As part of a Best Practices review of NSDI sites, looked node sites for the NSDI to get a sense of what is being done, what is working and valuable and what can be avoided. A report from this will be issued in the future.
- Reviewed and completed metadata for USGS products, DEM, DOQQ, DRG and the cropped collar DRG. Metadata for ALLARM, breeding bird atlas are currently under review. Metadata for Susquehanna River Basin Commission added as was metadata for the PA Fish & Boat Commission fish species atlas.
- Metadata management is in the process of being migrated into Spatial Metadata Management System software, a multi-user relational database solution for metadata is in the process. This will allow for simplified metadata creation by PASDA staff, more regular metadata "record" maintenance and importing of compliant metadata received from data sources.
- Completed metadata training for new staff
- Documentation completed for Alliance for Aquatic Resources Management (ALLARM) water quality data, 1995 SWP watersheds, PA GAP analysis Forest cover and support coverages.
- Metadata under review for GAP Managed Lands, PA Game Commission/GAP Breeding Bird Atlas and Lancaster County Metadata.
- Metadata in process for GAP PA Vegetative land cover, National Wetlands Inventory quad level data and Delaware River Basin HUC8 data. Ryan Baxter working with PSU Dept. of Geography to document Susquehanna River Basin level data.
- Online metadata submission form developed. Currently under testing in draft form with The Southwestern Pennsylvania Commission and Municipality of Murrysville. The form will guide the data contact through the process of metadata creation and submit metadata to PASDA/FGDC standards.
- Comments from Lancaster County GIS in their initial review of their metadata will be incorporated into a database (SMMS vendor software) which

will be maintained locally at Lancaster County. Complete PASDA/FGDC
metadata will be exported for inclusion in PASDA.

- PA GAP metadata updated.
- Managed lands in Pennsylvania updated.
- 100 hectare forested areas in PA created.
  Metadata for National Elevation Database (NED) county files created
  Updated coverage for PADWIS ground and surface water layers
- DRASTIC groundwater vulnerability scores
- Metadata for 1995 version of State Water Plan 104 watersheds
- Lancaster County (45 various layers)
- Stone Valley Experimental Forest layers
- National Park Service - Delaware Water Gap (5 layers)
- FGDC Metadata for the 1999 Cartographic/GIS Information CD-ROM is in
  process. Data files from the CD are being zipped for distribution will be
  added to PASDA with the metadata by mid-October.
- Susquehanna River Basin data from the Department of Geography.
- Data from the National Park Service at the Delaware Water Gap National Recreation Area.
- Stone Valley Experimental Forest data from the Forestry Department at Penn State.
- Data from the Pennsylvania Department of Transportation's CDROM.
- Created customized an Arc/Info AML to facilitate and automate the extraction of metadata
  information from any number of data sets at a time, rather than working with each data set
  individually.
- As a part of the FGDC Cooperative Agreements Program grant, a metadata
  training program is being developed. The half-day to full-day sessions
  will be focused on the FGDC metadata standard and the content of the
  metadata and will be able to adapt to include specific metadata tools that
  are in use. The first of these regional sessions will be sponsored by the
  Southwestern Pennsylvania Commission is tentatively planned for mid to late
  November. Plans for other sessions are being discussed with the City of
  Philadelphia and DVRPC.
- Flight dates for the Pa. DOQQ's have been obtained from the USGS. This
  listing will be posted shortly to the website while the dates are put into
  the metadata.
- 215 metadata and data files for boundaries and roads by county in ESRI shapefile and
  Geomedia warehouse formats as well as PDF printable/viewable Type 10 highway maps for
  PA counties.
- Moved grid data about the Stone Valley Experimental Forest from their server to PASDA and
  created metadata for it. Organized all of the Stone Valley Experimental Forest data (grid,
  vector, and ArcView legend files) for packaging and transfer to the PASDA server.
- A training session has been scheduled sponsored by the Southwestern Pennsylvania
  Commission and with cooperation of Three Rivers Connect, URISA and the University of
  Pittsburgh. The half-day session will be held at the SPC offices in Pittsburgh on December 9.
- Discussions are in process with the Philadelphia Mayors Office of Information Service and
  DVRPC about a similar session in Southeastern PA.

Meetings and contacts:

Met with Kevin Maxfield (Union County), Stewart Bruce (Mifflin County), Dale Bruns (Wilkes
Univ.) and Eric Jesperson (PaMAGIC) on October 20. The meeting was intended to coordinate
activities and knowledge regarding the FGDC 'Don't Duck Metadata' Cooperative Agreements
Program. The goals of each organization involved were outlined and while no formal cooperation
was agreed to, continued communication will find areas for partnership and cooperation. Wilkes
will be focusing training activities supported by their grant in Northern and Northeastern
Pennsylvania.

Mike Furlough, the Associate director of the Geospatial and Statistical Data Center of the University of Virginia, visited in conjunction with the WebGIS conference held at PSU. He is a part of a taskforce designing a clearinghouse of geospatial and non-geospatial data for UVA and wanted to share information and experiences with data documentation and dissemination through PASDA.

Discussed updates to SEDA-COG metadata with Joseph DeWalle and Tom Bresenhan. Changes to existing data and new data documentation will be reflected in PASDA.

Chris Rearick of the Municipality of Murrysville is working with a prototype of a metadata submission form. As a result of this testing, some changes are being incorporated in the form.