User Needs
&
Operational Requirements for
the Pennsylvania
Geospatial Data Clearinghouse

Pennsylvania Spatial Data Access
(**PASDA**)

August 10, 2000

Table of Contents

EXECUTIVE SUMMARY

BACKGROUND

Pennsylvania Spatial Data Access, PASDA, was named the official geospatial data clearinghouse for the Commonwealth of Pennsylvania on September 9th, 1999. PASDA had been serving as a node for Pennsylvania on the National Spatial Data Infrastructure (NSDI) since 1996 through funding from the Pennsylvania Department of Environmental Protection (PADEP), and providing metadata and free data for an array of state and local government agencies. Discussion related to implementing an official clearinghouse for the state, which would serve as the official repository of state GIS data and metadata, began in July 1999. Based on these discussions, the Services Committee of the Pennsylvania Geospatial Information Council, PAGIC, the body that oversees the operations of the clearinghouse, decided to examine the user requirements and state-wide needs related to the functions and services of the clearinghouse. The results of this examination or needs analysis will guide future development of the official clearinghouse.

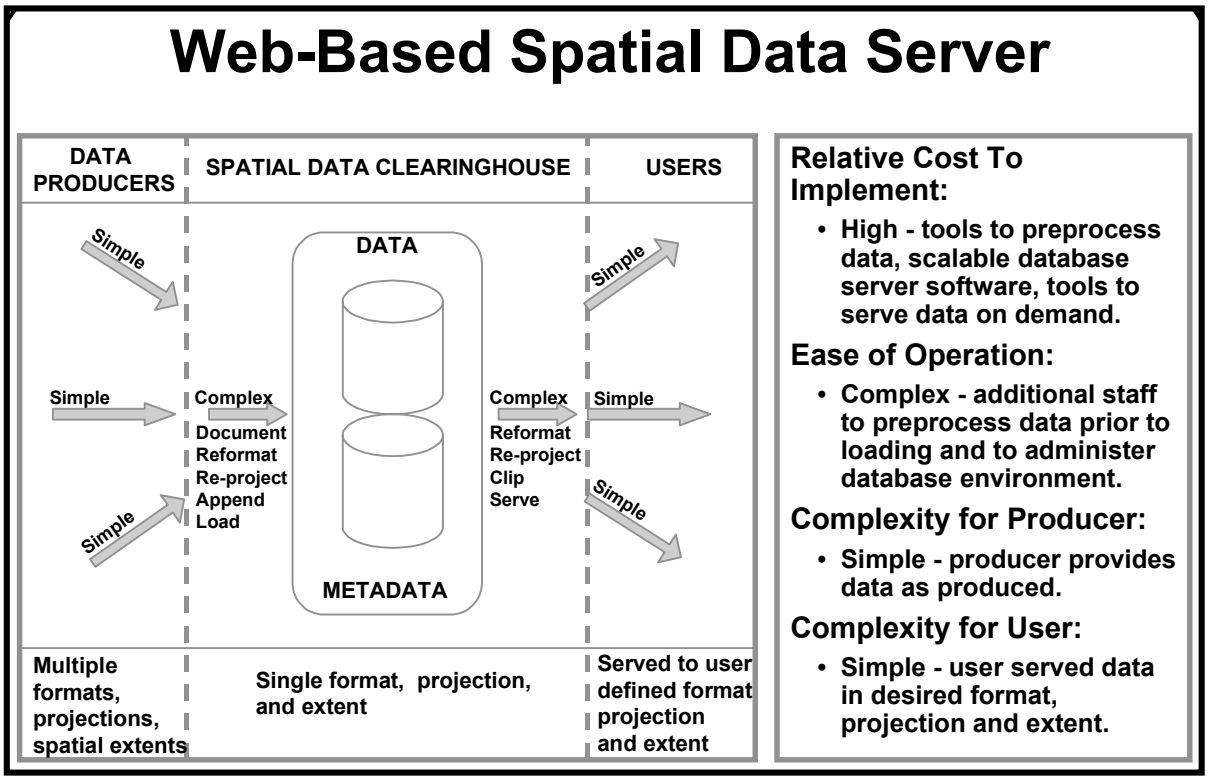
REQUIREMENTS

Functional—The following list reflects the most important elements of the clearinghouse as stated by the users:

- Ease of use
- Advanced data manipulation, viewing, and download capabilities
- Data accessibility
- Quality metadata and metadata input/training resources
- Effectiveness of on-line assistance/education
- Responsiveness to data providers
- Responsiveness to data users
- Responsiveness to PAGIC members

From the input received during the needs analysis process, it is clear that PASDA users want increased functionality in the clearinghouse. This functionality includes the development of user defined searches and data downloads, expanded web GIS capabilities that allow customized selection, viewing, and display of available data, and more advanced options such as clipping, reprojection, and analysis capabilities. When users are presented with various configurations for the clearinghouse functional architecture, they invariably choose the architecture that places the burden of effort on the clearinghouse versus the user. This architecture which requires the development of a relational database and accompanying functionality, should also be simple to use and easily accessible. These functions will allow a user of any level of ability to easily access, manipulate, and download data from the PASDA clearinghouse. It will also provide more advanced users with the ability to analyze data sets or to access remotely held data.

In order to achieve these, a new architecture will need to be developed for the clearinghouse. The current structure, which consists of ftping zipped files, will have to evolve into a more complex and adaptable web based spatial data design. The figure below illustrates the desired architecture:



Data—PASDA users demand access to a wide variety of data, at many scales, from many providers. The most requested data will fall within the seven framework data themes:

- Geodetic Control
- Orthoimagery
- Elevation
- Transportation
- Hydrography
- Governmental Units
- Cadastral Information.

However, there is a demonstrated need for data that falls outside of these categories. Uses of PASDA data ranged from emergency management to biological research, as well as land use planning, farmland preservation, and demographic analysis. It is also highly important for the clearinghouse to provide access to local and regional government data as well as state-wide data.

Services—The expectations for clearinghouse services were wide ranging and reflected functional and data needs. The expected services of the clearinghouse are:

- Access to data and metadata either centralized or remotely held
- Metadata development and training
- Online metadata submission, templates, and tools
- Data manipulation capabilities via the web site
- Data contacts or liaisons
- Online tutorials

First and foremost, the clearinghouse should serve as the primary vehicle to access geospatial data in the state. Access to the data can be provided in several ways—through direct download, links to data providers through metadata, and through advanced data manipulation capabilities via the web. In addition, PASDA should provide training opportunities in relation to metadata development and use of PASDA data and the web site, online tutorials and instructions, online metadata submission capabilities, and outreach to K-12 institutions. In addition, PASDA must develop and formalize data agreements, contact or data liaisons, and data acquisition processes that enable data providers to easily provide, update, and access their data and metadata.

IMPACT

The impact of this needs analysis and continuing discussions and guidance from PAGIC will be increasingly improved services offered to an expanded customer base, flexibility for on-line use of data as well as customizable download options for on-line data, and expanded promotional and partnership activities. This analysis combines input from GIS practitioners, government stakeholders, and non-profit organizations and uses this input as the basis for the creation of services, the acquisition of data, and the development of outreach services. The goals of this analysis are to develop the PASDA clearinghouse into a system that assists state and local governments in better serving the citizenry of the Commonwealth through cost savings and increased communication and cooperation.

USER NEEDS & OPERATIONAL REQUIREMENTS FOR THE PENNSYLVANIA GEOSPATIAL DATA CLEARINGHOUSE, PENNSYLVANIA SPATIAL DATA ACCESS (PASDA)

1.0 INTRODUCTION

BACKGROUND LEADING TO THE STUDY

Pennsylvania Spatial Data Access, PASDA, was named the official geospatial data clearinghouse for the Commonwealth of Pennsylvania on September 9th, 1999 by the Pennsylvania Geospatial Information Council, PAGIC. Discussions related to this designation began in April 1999 through PAGIC, the body that oversees the operations of the clearinghouse. In July 1999 the Services Committee of the PAGIC decided to examine the user requirements and statewide needs related to the functions and services of the clearinghouse. The PASDA staff, along with consultants from Advanced Technology Solutions and Soza & Company, formulated a methodology, time-line, and tasks to complete the needs analysis. Throughout the period from August 1999 to June 2000, the needs analysis team conducted research, compiled data, and held user input sessions, which serve as the basis for this report.

PURPOSE OF THE STUDY

The purpose of this study is to identify the user functional and data requirements for the official Pennsylvania Geospatial Data Clearinghouse.[1] This study will guide future development in order to remain responsive to the needs of PAGIC and PASDA users.

METHODOLOGY

The project consisted of three phases:

- Scoping
- Data Gathering
- Analysis

During the Scoping phase, the Core Team (consisting of PASDA representatives, as well as consultants) mapped out the approach to assess the user requirements and operational impacts resulting from the needs analysis project. Staff were assigned specific tasks and responsibilities related to the study.

[1] For the purposes of this document, the terms Pennsylvania Spatial Data Access, PASDA, the Pennsylvania Geospatial Data Clearinghouse, and "the clearinghouse" all refer to PASDA.

The data gathering phase consisted of two stages. The first required the project team to develop an extensive online web survey consisting of a key set of questions related to the desired functions and services of PASDA. Users were provided with an opportunity to supply input, comments, and suggestions via the PASDA web site. This survey was open to all PASDA users. The second phase used the data gathered through the web survey to identify and select users who would attend one of three user input sessions. A more extensive survey was developed and individuals were selected to attend one of three Group Systems sessions (Group Systems is a software program which allows collaborative decision making). The project methodology required that these participants come from a wide variety of organizations. One session was held, in Harrisburg, specifically for state government agencies in the Commonwealth. This session was followed by two additional sessions—one for local and regional government, non-profit, and business representatives, and another specifically for emergency management personnel, which were held at the Penn Stater Conference Center in State College. The format of these sessions was based on the format of the PASDA annual review, which is a regularly held meeting to evaluate the effectiveness of the PASDA system.

The Analysis phase, of which this report is a result, entailed an examination of user responses to the web survey and Group Systems sessions and the development of recommendations based on these responses, experienced project staff, and the research conducted in the scoping section of the project.

2.0 PURPOSE OF THE CLEARINGHOUSE

VISION FOR THE CLEARINGHOUSE

The vision for the clearinghouse provides guidance and direction for all initiatives undertaken by the clearinghouse management and staff. It also serves to state, in broad terms, the overall goal for the project:

> AS THE OFFICIAL STATE GEOSPATIAL DATA CLEARINGHOUSE FOR PENNSYLVANIA, PASDA WILL BE THE PREFERRED AND MOST COMPREHENSIVE SOURCE FOR LOCATING AND ACCESSING ALL AVAILABLE DIGITAL SPATIAL DATA FOR THE COMMONWEALTH. PASDA WILL BE THE PORTAL TO BOTH CENTRALIZED AND DISTRIBUTED DATA AND, IN DOING SO, WILL BE RECOGNIZED FOR THE COST SAVINGS AND SERVICES IT PROVIDES TO THE GOVERNMENTS, CITIZENS, AND BUSINESSES OF THE COMMONWEALTH.

In order to achieve this vision, PASDA has developed a series of statements that serve as guiding principles for the project. Critical to the success of the project, the following statements serve as the guiding principles for project goal setting and decision-making:

- PASDA will strive for open communication and establish cooperative partnerships in it's dealings with PAGIC, citizens, organizations, governments, and businesses within the state in order to promote the sharing of ideas, data, and tools related to geospatial information.

- Users have access to a comprehensive set of spatial data and associated metadata comprised of most current, historical, and time series data. Users have free public access to on-line data, and data that are downloadable are provided "free of charge" to the public.

- PASDA will provide training and outreach services to the users and data providers through presentations and training workshops focusing on developing and maintaining metadata, using Clearinghouse functions, and using the data available on the Clearinghouse.

- PASDA will develop a clearinghouse for educational resources related to GIS and spatial data for K-12 students and teachers and novice users offering online educational tools and links to tools which enable them to access, use, and understand GIS applications and the uses of spatial data.

- PASDA provides access to high quality metadata through its comprehensive metadata catalog. Through this catalog, users will have access to contact information for data not available for download from PASDA including metadata for fee-based data available sources other than PASDA such as public and private producers.

- PASDA will be flexible and responsive in adapting to new and urgent needs as well as new opportunities. The PASDA site will easily incorporate new data, ideas, and functionality.

- PASDA will continue to monitor technological developments and, whenever practical, desirable, and cost effective, will promote innovation and incorporation of new technology.

- PASDA will strive to be responsive to all requests for information and action, and will establish deadlines and guidelines for timely feedback and follow-up.

FACTORS RELATED TO MEASURES OF CLEARINGHOUSE PERFORMANCE

An important question is "What are the key characteristics that define the success of PASDA in fulfilling its Vision, and how do we measure our progress toward achieving the desired vision?" To this end, the following key characteristics should periodically be evaluated to determine the progress toward the desired Vision:

- Ease of use
- Data accessibility
- Quality metadata
- Effectiveness of on-line assistance/education
- Responsiveness to data providers
- Responsiveness to data users
- Responsiveness to the PAGIC members

A variety of descriptive measures and associated metrics should be developed by PASDA and tracked over time in order to measure progress in achieving these key characteristics.

3.0 USER FUNCTIONAL REQUIREMENTS

DATA AND METADATA SEARCHING AND BROWSING

Providing users with an easy way to locate the data sets they require is one of the most important services a clearinghouse can provide. This function deals specifically with data discovery and not with data viewing. The questions relating to this function were designed to determine:

- Whether metadata searches could be made easier,
- The best ways to improve the metadata search process,
- Whether the data catalog system for organizing data holdings should be improved, and
- The best way to make improvements to the data catalog.

3.1.1 The Present PASDA Approach

3.1.1.1 Data Searching and Browsing

Data holdings can be browsed without using the metadata by looking through the data catalog. The data catalog is an alphabetized list of data holdings. In addition it is possible to browse the FTP site—ftp://www.pasda.psu.edu/pub/pasda/—directly.

3.1.1.2 Metadata Searching and Browsing

Presently there are three paths to access PASDA's metadata holdings:

- The data catalog(located at http://www.pasda.psu.edu/ftp/ftp.html.)
- The text-based metadata search facility enables keyword Boolean searching of all fields in the PASDA metadata. located at http://www.pasda.psu.edu/search/search.html.)
- A spatial metadata search capability that searches all metadata records by county, drainage basin, or USGS quadrangle boundary. It is also possible to combine such a spatial search with a keyword search (located at http://www.pasda.psu.edu/search/map_search.html.)

3.1.2 Analysis of the Survey Results

1) How to make metadata searches easier:
   - Users said that searching was generally good.
   - The text-based search capability should be made easier to use.

2) The best ways to improve the metadata search process:

   - For metadata search approaches, geographic (using a spatial extent as part of the metadata search) was most often chosen as the preferred way to search, followed closely by theme, and then by publisher, accuracy, and date of last update/currency. Some said the present geographic option of searching by quads is not easy for those not familiar with USGS quadrangles.

   - Find a way to eliminate multiple responses of the same data type.

3) Whether the data catalog system for organizing data holdings should be improved:

   - Several people said that the combination of provider and data set listings was confusing .

4) The best way to make improvements to the data catalog:

   - There needs to be a better organization system for cataloging the data sets, making it easy to find data by data type as well as by provider .

3.1.3 Recommendations

3.1.3.1 Data and Metadata Searching and Browsing

Recommendation 1—Order the Data Catalog Separately by Theme and by Provider[2]

The present data catalog should be reworked into two separate catalogs so that users can choose to search lists organized by either the type of data—geology, land use, etc.—or by the data provider.

1) Catalog Organized by Theme—The theme list should be organized based on the standard hierarchical system of organizing spatial data. Some of the broad categories in such a system might be:

- Physical Environment—geology, soils, contours, DEMs, hydrology, watersheds.
- Natural Environment—land cover, habitats, endangered species, wetlands.
- Natural Resources—forestry, energy, minerals, agricultural, game management.
- Meteorology—temperatures, winds, precipitation, growing seasons.
- Images—satellite, aerial photo, maps.
- Infrastructure—transportation, utility, structures.

[2] This Recommendation responds to needs defined in Parts 3 and 4 of the Analysis section above.

- Cultural Resources—archaeological, historical, land use, recreation, industry, arts.
- Cultural Boundaries—national, state, county, municipal, agency, school, electoral.
- Census—demographic, economic, census boundaries.

2) Catalog Organized by Data Provider—This listing would be based on the Data Provider Table (as defined in Recommendation 2 below), and would allow users to search an alphabetized list of providers.

Recommendation 2—Create a Data Provider Table[3]

Users said a listing by data provider would be useful. In order to address this need a provider table should be created that identifies data providers with data on PASDA. There may be a benefit in also using this table to track other spatial data located in Pennsylvania, including those data not residing on the PASDA server. Such a table would provide a comprehensive link between data users and all of Pennsylvania's data resources, and would address the need many expressed in the workshops that PASDA be a clearinghouse for all data in the Commonwealth, in addition to state agency data

This table would contain a primary key that would relate to the metadata for each provider. Minimum criteria for inclusion on the provider list might be:

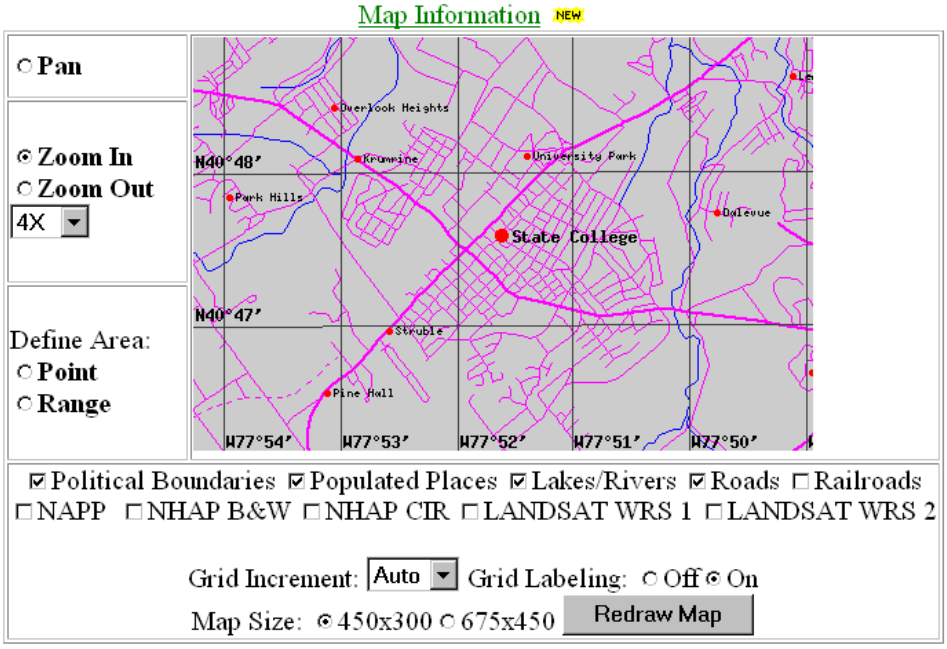
- Organization Name,
- Contact Information (email, phone, and/or mailing address),
- Access type (public domain, fee, etc.),
- Access protocol (on-line FTP or HTTP, mail order, etc.),
- Data theme(s) (list major themes),
- Map extent of data (approximate),
- Projection(s) space of the data (list main projection spaces used),
- Data scale(s) (list major scales used), and
- Metadata availability.

Recommendation 3—Map Based Tool to Define Spatial Extent for Metadata Search[4]

A tool should be implemented which allows a user to interactively define a map extent for a metadata search. All metadata records falling within the map extent could be retrieved. An example of such a search tool is on the USGS Global Land Information System (GLIS) site at http://edc.usgs.gov/webglis. This tool can specify a geographic area, as well as text based searches on key fields. (See figure on next page)

[3] This recommendation is designed to address needs defined in Parts 1, 3, and 4 of the Analysis section above.

[4] This recommendation is designed to address needs defined in Parts 1 and 2 of the Analysis section above.



For a geographic search engine to work the minimum and maximum (x, y) coordinate fields should be populated in the FGDC metadata.

Recommendation 4—For tiled layers, only display one metadata record on query[5]

For those metadata records associated with spatially-partitioned or tiled data, there should be a master record which corresponds to the theme returned in a query. From this record the user can link to the full list of metadata available for each tile, if desired. This will be necessary if PASDA is to move data into a seamless and continuous layer stored within a central data server as identified in Section 5.4.

Recommendation 5—Combined Spatial and Text-Based Metadata Searching[6]

Geographic coordinates should be considered another criteria in a search environment that combines geographic and text based options. For users familiar with FGDC metadata fields and Boolean syntax the capability to search multiple metadata fields should be provided.

Recommendation 6—Provide Easier Access to the FTP site[7]

FTP access could be provided as a heading directly under the "Search" heading on the home page.

[5] This recommendation is a response to Part 2 of the Analysis section above.

[6] This recommendation is designed to address needs defined in Parts 1 and 2 of the Analysis section above.

[7] This recommendation is designed to address a need defined in Parts 1 and 2 of the Analysis section above.

Recommendation 7—Provide an Abbreviated Metadata View[8]

There is already an abbreviated metadata view provided. However, in some cases it may be too abbreviated. In order to allow easier interaction with the metadata, there should be a default metadata view that contains the most essential fields with an easy-to-understand field definition. Such a metadata view might include a select list of fields with easy-to-understand versions of the field names, such as:

Who provided the data to PASDA?	PennDOT
What kind of data is stored in this file?	State road data
At what scale is this data meant to be used?	1:24,000
When was the data last updated?	10/10/99
What is the projection?	Polyconic
In what datum is the data in?	NAD83

Recommendation 8 – Provide a Customizable Metadata View

Because of the difficulty of having a "one-size-fits-all" metadata format, a user-customizable metadata field view would be valuable. Such a capability would allow the user to choose any fields from the FGDC format. This type of capability may make sense as one of several capabilities that custom profiles could provide. See Section 3.6 below for further discussion of user profiles.

INTERACTIVE MAPPING AND DATA PREVIEW OPTIONS[9]

To simplify the users' ability to determine what data they want, and perhaps avoid the need to download data, online data mapping and viewing could be provided. Once an environment exists which allows users to view data, other analytical and educational opportunities could present themselves, including applications geared to teaching that might use a variety of geographically-referenced digital media.

PASDA is housed within an educational institution. In this context, there are services (in addition to those relating purely to data access) which should be contemplated. One service area falls into the broad category of "subject-based services." These are services tailored to a subject area, providing opportunities for in-depth learning in that subject and possibly making use of a variety of media types. In this light, the questions relating to this function were designed to:

1.) Determine whether users feel a need to preview data before download,

2.) Determine what the desired level of interactive mapping is for a clearinghouse, and

3.) Determine what subject-based interactive mapping applications would be the most valuable at PASDA.

[8] This recommendation is designed to address needs defined in Parts 1 and 2 of the Analysis section above.

[9] Web-based GIS have many dimensions: interactive mapping, reprojections, reformatting, and clipping, to name a few.

3.1.4 The Present PASDA Approach

Presently, the only data that can be interactively browsed are the MrSID versions of the DOQs. Static preview images of data sets are available for most downloadable datasets on PASDA.

3.1.5 Analysis of the Survey Results

1) Determine whether users feel a need to preview data sets before download:

- Most respondents said they would regularly use a tool that would allow them to preview data prior to downloading it, if it were available.

2) Determine what the desired level of interactive mapping is for a clearinghouse:

- Most users wanted a level of interactive mapping beyond the ability to pan and zoom with a standard base map backdrop; many requested the ability to identify attributes of features and build thematic maps.

3) Determine what subject-based interactive mapping applications would be the most valuable through PASDA:

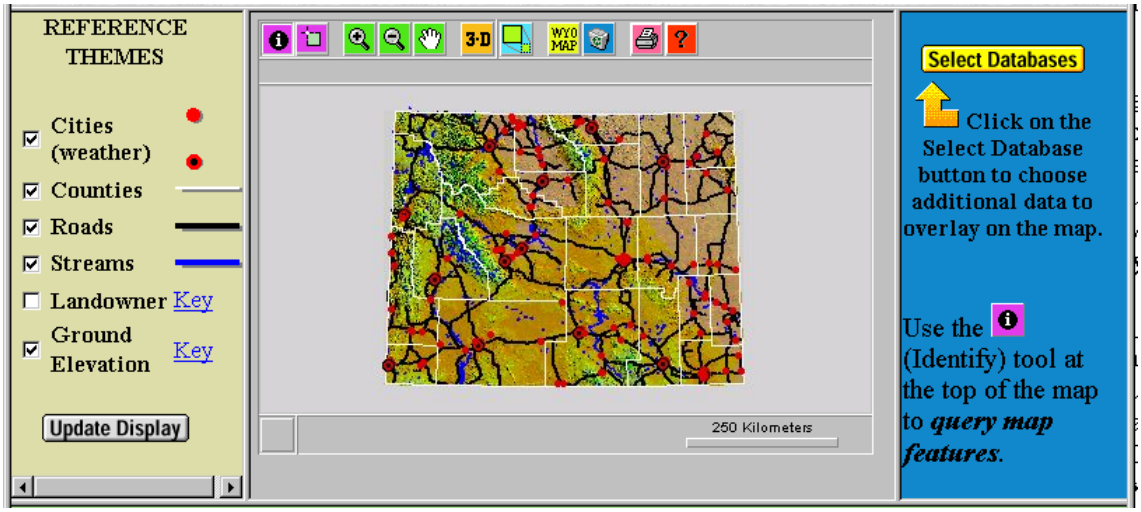
- The votes for applications relevant for the clearinghouse were as follows:

Application	# of Responses
Atlas of Pennsylvania	41
GIS Educational Resources in Pennsylvania	15
National Geodetic Monument Framework	12

3.1.6 Recommendations

Recommendation 1—Create An Easy-to-Use Data Browsing Environment[10]

"Ease of use," and "data accessibility," were seen as the most important characteristic for PASDA to implement an interactive mapping application that can also serve as a data preview option to view a particular data set prior to downloading the data. The data-browsing environment allows the user the option to view the level of feature detail (graphic resolution) and the classification scheme of a data set to determine if the data will fulfill a particular need. An interactive data-browsing environment should be developed for non-expert users. There is a good example of such a capability on the Wyoming Internet Map server at http://wims.sdvc.uwyo.edu/wyoims2/wims2a.html, which provides interactive browsing capabilities with predefined symbolization for a range of statewide data sets.



The interactive on-line data set previewing option should be limited to data that are loaded into a web-enabled spatial data server in a standard projection (Section 5.4). For "non-standardized" data sets stored as system files on a file server (similar to current practice), PASDA should continue to create static preview images of data sets. The static images should clearly show the resolution and

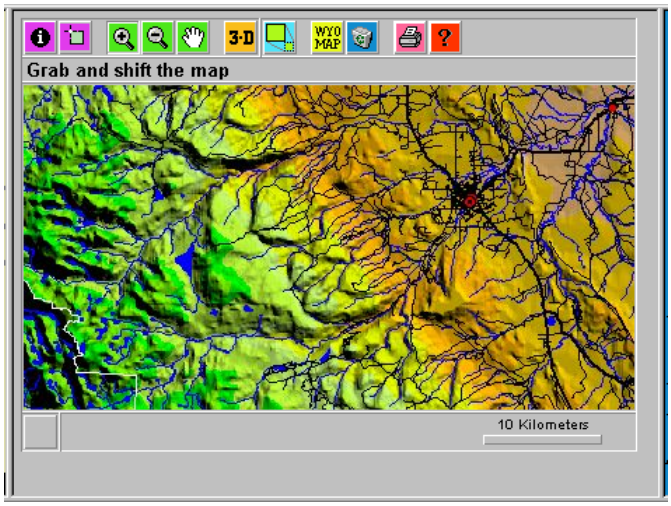
[10] This recommendation addresses needs defined in all three parts of the Analysis section above. The relevance for subject-based mapping is that such an easy-to-use viewing environment will be the base for subject-based applications.

density of the data set, and should be displayed on-line when users request data previews for "non-standardized" data Recommendation 2—Ability to Browse Data from the Metadata View[11]

Provide a "one-click preview" of the actual data, linked from the metadata itself. Such a feature would load the data browser next to the (abbreviated) metadata view (see Recommendation 5 in Section 3.1 above for discussion on an abbreviated metadata view), load the default base map, and then load the requested data set using predefined symbology.

Recommendation 3—Develop an Attractive Base Map[12]

A base map is an important part of an interactive mapping application because it is displayed as a supporting background almost any time data are viewed and provides geographic context. Such a base map might include state and county boundaries, major drainage, major roads, and elevation data that would be displayed in the background as a default whenever someone wanted to preview a data set. A key part of the base map, given the topography of Pennsylvania, is elevation. Therefore it might make sense to develop a shade relief raster image for the entire state. The Wyoming site has done this to excellent effect.



[11] This recommendation addresses the need defined in Part 1 in the Analysis section above.

[12] This recommendation addresses the need defined in Parts 1 and 2 of the analysis section above.

Recommendation 4—Keep Option Open to Distributed Data Servers[13]

It may be possible to allow browsing of data sets that are not on PASDA's server. This option should be kept in mind during the system design. Such a distributed model would address some of the hardware, processing, and manpower scalability issues that PASDA will face as the quantity of data it provides grows and as the number of providers and users increases.

Recommendation 5—Develop Subject-Intensive Interactive Mapping Themes[14]

Implementation of a web-enabled spatial data server architecture (Section 5.4) offers new opportunities for PASDA to develop subject-rich interactive mapping applications. The most popular was the "Atlas of Pennsylvania," which could have a variety of views for the state including geology, soils, landforms, land use, etc. The interactive mapping might be enhanced by adding multimedia components, such as clicking on historic resource locations to view digital pictures of historic structures or clicking on Valley Forge to see a video clip of the encampment.

DATA AND METADATA RELATED SERVICES

The services referred to in this section are of major importance to the overall design PASDA chooses and will be critical for enabling a web-enabled spatial data server (Section 5.4). These are the services PASDA will have to perform if it is to receive data in multiple formats, spatial extents, and map projections from data providers, while also allowing users to customize the desired map extents, formats, and map projections prior to download. The questions relating to this function were designed to determine:

1) The most valuable options for customizing data prior to download.

2) Specific projections, datums, and formats presently in use among users and data providers.

3.1.7 Present PASDA Approach

Presently PASDA receives data from many organizations in many formats and projections, and stores the data as system files on a file server, in most cases without modifying the data. PASDA generates the metadata for the data sets and makes the metadata available for searching and download with the data sets. It is possible to view the metadata in both a highly abbreviated form and a more complete format.

Because the user typically downloads data sets in a variety of projections and data formats, PASDA expends considerable staff time providing follow-up guidance to non-technical users who download the data. The follow-up support ranges from how select a suitable map projection for a

[13] This recommendation does not follow directly from the Analysis section, but is included here because of the potential long term value of integrating distributed, as well as centralized data, in an easy-to-use data browsing environment.

[14] This recommendation addresses the need defined in Part 3 in the Analysis section above and complements furthers the role of Penn State, as a educational institution, to develop educational resources and outreach activities.

particular project, how to project all the data into a common map projection, or how to load data that were downloaded in a variety of formats.

3.1.8 Analysis of the Survey Results

1) The most valuable options for customizing data prior to download:

- When asked whether they would like PASDA to add data services that would reduce work for them, people are interested. There is definite interest in having PASDA provide a range of on-line data customization options including re-projection, format conversion, and data set clipping to define subset areas of interest.

- If online data viewing is a service, then there is real interest in that capability as well. "Ease of use" is the top of the list for all groups. Clearly, if users can get answers to questions by viewing the data online and not having to download or reformat, that would be desirable.

2) The specific file formats, projections, and datum spaces desired for download.

The tabulations below summarize file download preferences.

Vector Data	# of Responses	Raster Data	# of Responses
ESRI Shape File Format	16	GeoTIFF/TIFF with World File	14
ESRI Coverage Format	6	JPEG Format	6
ESRI Export (.e00) Format	4	MrSID Format	3
AutoCAD (.dwg) Format	2	BIL Format	3
Microstation (.dgn) Format	2		

- The following table displays the survey results for GIS platforms used by organizations:

GIS Software	# of Responses
ArcView	15
Arc/Info	10
MGE	3
AutoCAD	3
GeoMedia	2
Microstation	1
Erdas	1
Ermapper	1
MrSID	1

3.1.9 Recommendations

Recommendation 1—Provide download options that can be implemented using automated techniques[15]

Consideration should be given as to how the implementation of automated service will affect the overall performance of the clearinghouse. Below are the services desired (listed in order by the number of votes they received):

1.) Custom Spatial Clip—Allow users to clip out a subset of the data before downloading. This is perhaps the most valuable potential service, given the modem speed of many users. To be implemented well, a wide range of polygon boundaries should be available as clip boundaries. Users should also be given the choice to interactively define a rectangular box in order to create a rectangular subset of data.

2.) Custom Coordinate Space Conversion—Allow users to specify a new projection space for the data before downloading. This will make data easier to use for both technical and non-technical users.

3.) Custom File Format Conversion—Allow users to specify a new format for the data before downloading. This will also remove a barrier for many potential spatial data users. While it might not be possible to support all file formats, the most widely used formats should be supported.

DATA AND METADATA ACQUISITION AND DISTRIBUTION

This function is concerned with enabling PASDA to get data from providers and for PASDA to deliver data to users. In some cases data can be sent electronically, but in other cases data must be copied to some medium and sent via mail. The questions in this section are therefore designed to determine:

1) The type of Internet connection users have,

2) The maximum acceptable time for downloads,

3) The best media types for delivery and distribution,

4) Whether metadata should be automatically included with data,

[15] This recommendation is designed to meet the needs defined in Parts 1 and 2 in the Analysis section above. These services are an important part of making PASDA easy to use. "Ease of use" and "data accessibility" were seen as the most important characteristic for PASDA to implement (Appendices A, B, and C).

5) Whether PASDA should make it easier for groups to create their own metadata, and

6) Whether metadata should be provided in both XML and ASCII formats.

3.1.10 Present PASDA Approach

3.1.10.1 Data

Presently PASDA acquires most data from federal and state agencies on a regular basis and then makes the data available for downloading from its server. In general, PASDA does not modify the data it receives but it creates metadata for most data it makes available.

PASDA accepts data in any format and projection. If the data are provided in spatial partitions or tiles, they are uploaded as separate files.

PASDA has created a Standard Data section that includes statewide data sets in decimal degrees and UTM.

PASDA has an agreement with the Pennsylvania State Data Center to provide PASDA data on CD for a fee.

3.1.10.2 Metadata

PASDA generates metadata for most of the data it makes available, and PASDA makes the metadata available for searching and downloading in XML format. When PASDA receives new data or data updates, it updates the metadata. On occasion, PASDA will make the data available for downloading before the metadata is available, typically because no metadata is available for the data.

3.1.11 Analysis of the Survey Results

1) The type of Internet connection users have:

   - About half of those questioned had a 56k modem or slower. This emphasizes the importance of the clipping service, as well as online viewing capabilities.

2) The maximum acceptable time for downloads:

   - Users would like to spend as little time as possible downloading data, with one to two hours seen as the maximum acceptable time.

3) The best media types for delivery and distribution:

   - For data that are too large to download, all groups desire an online request form.

   - CDROM was the preferred method of data exchange both for distribution and acquisition, followed by DVD and 8mm tape.

4) Whether metadata should be automatically included with data:

   - Respondents generally prefer a choice of whether to download metadata. When they select a file for download, they would prefer to be prompted whether they wish to append the associated metadata file to the download. One user indicated the metadata file name should match the data file name for ease of association when both files appear on their local hard disk, and another user suggested that at least some type of abbreviated metadata that documents the projection space should always be downloaded.

5) Whether PASDA should make it easier for groups to create their own metadata:

   - Presently very few groups maintain current metadata for the GIS data set they maintain.

   - People showed support for having PASDA make it easier for providers to develop FGDC-compliant metadata. This would be accomplished through training and online metadata creation capabilities. Whether metadata should be provided in both XML and ASCII formats:

   - People did not indicate any preference between metadata served as ASCII or XML format.

3.1.12 Recommendations

Recommendation 1—Remain Responsible for Metadata but Use Education to Encourage Data Providers to Become More Involved in Creating Metadata[16]

There are good reasons for providers to create and maintain FGDC-compliant metadata for their own convenience. It is simply good practice to keep track of one's own data, since the knowledge about data sources, lineage, quality, etc. are best known by those producing the data. By empowering data providers to create and maintain metadata, it will increase the likelihood that each data set on the clearinghouse will have accompanying metadata. This is especially important as the volume of data is likely to increase dramatically as more state agencies and local governments implement GIS. It is also in PASDA's interest to have data producers assume some of this responsibility over time, since PASDA can then focus on expanding outreach efforts. It also allows PASDA to focus on continuous enhancements to the clearinghouse. The following are ways for PASDA to encourage data producers to get involved with metadata:

- Choose a public domain (free) metadata tool and support it well,

- Explain the benefits of good metadata for the data distributor,

- Provide descriptions of GIS tools which automate metadata maintenance,

[16] This recommendation addresses needs expressed in Part 5 of the Analysis section above.

- Make an easy-to-follow tutorial available for metadata creation and maintenance using the tool of choice,

- Provide a Web-based metadata input form for those who would like to submit information about their data online, and

- Provide training in metadata creation and maintenance.

Recommendation 2—Provide Off-Line Delivery[17]

A useful service for those with low bandwidth connections would be to enable users, after selecting the data sets they desire, to request a CDROM or other media transfer instead of having to download the data. In fact, if the system determines that, for a given file size and connection rate, the time would exceed 1 to 2 hours, the system should automatically inform the user and provide the "offline download" request form. This service could be largely automated.

Recommendation 3--Define Requirements for Remote Storage of Data and Metadata[18]

The requirements for storage on remote servers should be defined so groups that wish to pursue that direction can do so. Such an architecture would treat data and metadata on remote servers as if they were local. Metadata might be copied at regular intervals to PASDA to create a unified metadata view. There could be advantages to all parties in exploring a "distributed data/central viewing" approach. It would decrease storage requirements. Additionally, the data provider would have less work replicating and sending data to PASDA. Data updates would reach the end user instantly.

Recommendation 4—Distribute Metadata in a Form that Loads Well to Metadata Tool[19]

There is a clear advantage to users and PASDA if downloadable metadata files can be easily loaded into a local metadata maintenance tool. Currently, nearly all tools will accept plain text files formatted using FGDC standard field names. These can also be maintained locally using a simple text editor. New tools are being developed that use Extensible Mark-up Language (XML) as a transfer format. So long as PASDA continues to store metadata in XML format, it can offer metadata downloads in a variety of formats according to user choice.

[17] This recommendation addresses the need defined in Parts 1 and 2 of the Analysis section above.

[18] This recommendation is defined to address PASDA's desire to provide broader access to spatial data in the state without making impossible demands on PASDA.

[19] This recommendation is designed to address a potential need in Part 6 of the Analysis section above.

ONLINE HELP AND TUTORIALS

The better the online help, the easier it will be for technical and non-technical people alike to use PASDA. In addition, being housed in an educational institution, it may make sense for PASDA to provide help geared towards educating those who are not knowledgeable about GIS and GIS data. The questions which relate to this function were therefore designed to determine:

1) The online help subjects or tutorials that would be most useful for using the site or accessing the data,

2) The most useful tools for providing online help, and

3) The topics which might be most valuable for helping elementary and high school students learn about GIS.

3.1.13 Present PASDA Approach

There is a Frequently Asked Questions (FAQ) Web page, extensive online tutorials and a mailing list and personal assistance through email. The FAQ includes detailed responses to questions and the entries found in the FAQ could be considered tutorials. The online tutorials include lessons on downloading data, importing and using data in ArcView, and will soon cover other software programs and functions. The mailing list is not a list serve type of mailing list, but rather one which PASDA can use to inform users. Personal assistance is always available through the pasda@psu.edu email address.

3.1.14 Analysis of Survey Results

1) The online help subjects or tutorials that would be most useful for using the site or accessing the data:

Online Help Subjects or Tutorials	# of Responses
How to convert between formats	22
How to convert between projections	22
How to increase the likelihood of success using metadata searches	16
How to use public domain tools to create and maintain metadata	15

- Many other responses were given including "down load procedures," "metadata production," "data update cycle/policy," and "data integrity issues."

2) The most useful tools for providing online help:

Tool	# of Responses
Keyword Searches	34
FAQ	32
Available Help Topics	30
Online Discussion Forum	21

3) The topics which might be most valuable for assisting elementary and high school students learn about GIS. These were listed as free form suggestions and the following were mentioned more than once:

   - Uses of geographic data,
   - Selection and use of suitable map projections and scales,
   - Career-related information, and
   - Links to other sites which provide online GIS tutorials.

3.1.15 Recommendations

Recommendation 1—Make Full Use of the Proven Combination of FAQ, Listserve-Based Discussion Forum, Email Support, Site Map, and Site Search[20]

Using this approach, an effective constantly improving support capacity can be developed using limited resources. Some pieces are in place now. For example, there is already a FAQ. If implemented carefully each piece reinforces the other, resulting in a knowledge base that continues to expand, driven in part by the users themselves.

- FAQ—To be effective, the FAQ must be regularly updated with answers to those questions people are asking more than a couple times. There should be a text search capability on the FAQ to make it easier to find information.
- Listserve—Start a listserve and invite all PASDA users to join. This is perhaps one of the most important things PASDA can do to improve service and it does not take much effort. A listserve is a mailing list of all those who wish to subscribe. It makes it easy for people to post a question to the entire user community, and for anyone to respond. This could reduce PASDA's load and improve information sharing in the state. A powerful feature of a listserve is the ability to perform a text search of the listserve archive and to review discussions by thread (or topic). This makes the past listserve discussions valuable to the entire community for answering questions that would otherwise go to PASDA or go unanswered. (One decision PASDA will have to make is whether PASDA should moderate such a listserve. It would be better if the listserve were moderated, as it would ensure that someone at PASDA knew what was being discussed, and could respond in a timely way.)
- Site Search—It should be possible to perform a search of all the HTML (excluding the metadata) on the PASDA site. This is a standard web site feature.
- Site Map—A site map is a standard feature of complex web sites like PASDA.

The Site Search and Site Map functions should be available from the home page. The FAQ, Listserve, and Email support should be grouped together. It should be made clear that most

[20] This set of recommendations was developed to meet needs defined in Parts 1, 2, and 3 of the Analysis section above.

questions can be quickly answered by checking the FAQ and the listserve archives and that email support is available if the user is still unsatisfied.

Recommendation 2—Work With a Committee of Teachers to Define the Best Help Topics for Elementary and High School Students[21]

Tutorial topics, if they are implemented, need to be carefully targeted by age group. In order to be sure the tutorials are widely used, it is important to work closely with educators to avoid reinventing the wheel and to raise awareness about the available resources.

There is a close link between this recommendation, and the subjective-intensive interactive mapping recommendation discussed in Section 3.2 (Recommendation 5). Careful thought needs to be given as to whether the tutorial topic would be better implemented as an interactive map-based interface, or whether the topic is better implemented with a text-based interface with supporting interactive mapping functions. In either case, the tutorial topic could be fully implemented with multimedia components including sound, video clips, and/or digital pictures.

USER SUBSCRIPTION SERVICE OPTIONS

Allowing users to define profiles for themselves based on a login and password could make many activities on PASDA more convenient for the user. Questions relating to this function were therefore designed to determine the following:

1) Whether users would take advantage of a subscription-based service that automatically notifies them about data updates based on user-defined criteria,

2) The criteria that would be most valuable to specify in setting up a profile,

3) Whether the profile function should be expanded for use in defining metadata search or data download preferences, and

4) The preferred method for un-subscribing from such a service.

3.1.16 Present PASDA Approach

Presently there is no way for a user to define a personal user profile on PASDA.

3.1.17 Analysis of Survey Results

1) Whether users would subscribe to a service which notified them about data updates for data sets which fit criteria they defined:

    - Most people say they would subscribe to a service that would allow them to define a user profile and could be used to automatically notify them if modifications were made in data of interest that fits their profile.

[21] This topic recommendation addresses the need defined in Part 3 of the Analysis section above.

2) The criteria that would be most valuable to be able to specify in setting up a profile:

Below is the number of votes for each category:

Category	# of Responses
Geographic Area	33
By Data Category	30
By Date of Last Revision	27
By Data Producer	9

3) Whether users felt the profile could be valuable to employ during metadata searches or data downloads:

- Most said they would like a profile to store information that makes it more convenient to perform metadata searches and define download preferences.

4) The preferred method for un-subscribing from such a service:

- Most users preferred to have the ability to unsubscribe themselves while online.

3.1.18 Recommendations

Recommendation 1—Implement an optional profile based system for users[22]

Such a "Personal Clearinghouse" system would be based on personal accounts that users could set up and configure to their choosing. Useful criteria for customization, that would be easy to implement, could use any metadata field. Custom profiles could be used in many ways, including:

- Automatic email notification when data fitting profile criteria are changed or updated and

- Preset parameters that could be used to define and obtain data; parameters could define data reformatting and delivery options including: acceptable clip area, projection, datum, data format, and delivery method.

USER FEEDBACK OPTIONS

Having an efficient user feedback mechanism is important both to address specific user issues or technical problems and also as a way for PASDA to measure customer satisfaction over time. Customer feedback also provides input for planning future changes to PASDA services. The questions for this function therefore attempt to determine:

[22] This recommendation was based on needs defined in parts 1, 2, and 3 of the Analysis section above.

1) Preferred methods for reporting problems with downloads or connections,

2) Preferred approaches to arranging offline data distribution,

3) The maximum acceptable time a user would wait to have a reported problem acknowledged by PASDA, and

4) Preferred approaches to solicit feedback regarding customer satisfaction.

3.1.19 Present PASDA Approach

There is presently a feedback form that can be filled out on the Web and sent to PASDA. Users can also email PASDA, the PASDA webmaster or use the PASDA guest book.

3.1.20 Analysis of the Survey Results

1) Preferred methods for reporting problems with downloads or connections:

Method	# of Responses
Online Form	32
Email Address to Web Master	31
"1-800" Number	27

2) Preferred approaches to arranging offline data distribution:

Method	# of Responses
Online Form	31
Email Address to Web Master	26
"1-800" Number	26

3) The maximum acceptable time a user would wait to have a reported problem acknowledged by PASDA:

- Many different answers were given here ranging from "as soon as possible" to four hours to two weeks.

4) Preferred approaches to solicit feedback regarding customer satisfaction:

Approach	# of Responses
Provide online customer satisfaction survey	21
Require user to log in	19
Provide a customer satisfaction survey file	11
Provide optional sign-in page	9

3.1.21 Recommendations

Making it easy for people to provide feedback to PASDA is important for ensuring customer satisfaction. It should be pointed out that none of the approaches mentioned would result in a statistically valid measure of customer satisfaction. In order to achieve a thoroughly objective survey, PASDA would need to randomly contact users who have connected at least once to the PASDA site.

Providing a range of feedback avenues will allow people to choose the one with which they feel comfortable.

Recommendation 1—Survey

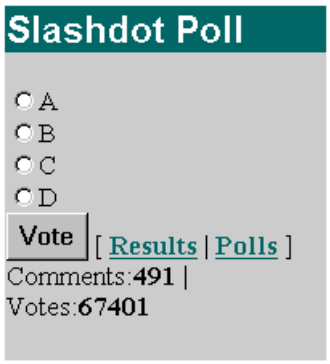
In addition to the present "Comments" form, a longer online survey should be made available for those who might want to use it. This should be mostly multiple choice with a number of questions requiring longer responses. No question should be mandatory.

Recommendation 2—Easy Email Contact With Web Master/Customer Support

Links to forms that make it easy to send email to support personnel is a must. The links should be available under a "Contact Us" heading. The web master link should be at the bottom of each page. A form is more useful than an email link because the email link requires that a mail tool be properly configured.

Recommendation 3—Online Poll

A mini "poll" may be useful. This could be placed on the home page on the left at the bottom of the form. A question could be, "How are we doing?" The answer check boxes might be, "Great Site," "Does the Job," "Needs Work," etc. By submitting a vote the user can see the month's vote tally. A response of "Great Site" would result in a text box requesting "What do you like best?". If "Needs Work" is chosen, a text box could come back with "Please tell us which areas need the most work, and give your suggestions for making it better." The image at the right is an example of how a poll can be implemented on a web page using very little space.



4.0 USER DATA REQUIREMENTS

PRIORITIZED NEEDS FOR SPATIAL DATA SETS

PASDA has a significant collection of federal and state agencies' data, but there are many data sets which exist within Pennsylvania that PASDA has not yet acquired. There are also data sets which users would like to have available for which they do not have access. One of PASDA's ongoing tasks is to keep its data holdings in line with user needs. The questions for this section were therefore designed to determine:

1) The priority data sets that need to be made available online and maintained as up-to-date as possible and

2) The data sets presently being maintained by organizations which are not on PASDA but which would make sense to be distributed by PASDA.

4.1.1 Present PASDA Approach

Presently PASDA accepts any Pennsylvania spatial data. There are data sets which users would like to have which no agency is presently producing.

4.1.2 Analysis of the Survey Results

1) The priority data sets that need to be made available online and maintained as up-to-date as possible:

   - Below is a list of the data sets and the number of times they were mentioned in response to the question, "Identify the top five GIS-compatible data sets maintained by external organizations that you either do not currently have access to or that you wish were more "up-to-date."

Data Set	# of Responses	Data Set	# of Responses
Digital Ortho Quadrangles	5	DLGs	1
Soils—STATSGO with Attributes	3	High Resolution Color IR Imagery	1
Digital Raster Graphics	3	PA Natural Diversity Index	1
Tax Parcels	3	Historic Maps	1
Census 2000	2	Zip Codes	1
Digital Elevation Models	2	Water Monitoring	1
Landsat TM Data	2	Geology	1
Floodplains	2	Roads with Street Names	1
Streams with Streams Labeled	2	National Wetlands Inventory	1
Landuse	1		

2) Data sets presently being maintained by organizations which are not on PASDA but would make sense to be distributed by PASDA:

   - Below is list of the data sets and the number of times they were mentioned in response to the question, "Identify the top five GIS-compatible data sets created and/or routinely maintained by your organization that would be the highest candidates for distribution by the clearinghouse:"

Data Set	# of Responses	Data Set	# of Responses
Street Network	3	Active/Nonactive Landfills	1
County Roads	3	Emergency Service Zones	1
Cooperative Multi-site Agreement Sites	2	Recreation Facilities	1
County Tax Parcels	2	Contour Lines	1
Hydrology	2	Archeology Sites	1
Watersheds	2	Harvest Data by County	1
State Game Lands	2	Landcover	1
Susquehanna River Basin Commission Water use Projects	2	County Land Use	1
Susquehanna Basin Polygon	2	DEMs	1
Historic Structures and Districts	2	DOQs	1
Natural Heritage Regions	1	Municipal Mapping	1
Lebanon County Geology	1	Municipal Boundaries	1
State Roads	1	Approximate locations of Archeological Sites	1
Geology	1	Right of Way	1
Mine Subsidence	1	Pennsylvania Natural Diversity Index Rare Species Data	1
Zoning	1	Buildings	1
Act 2 Land Recycling Sites	1	WCO Districts	1

Of those data sets in Part 1 above that users say they would like to have, many are mentioned in the list under Part 2. Interestingly, many of the data sets which match across the lists are those which are most valuable at the local level: tax parcels, land use data, roads with street names, and detailed soils. These data sets are generated locally but are valuable to others in the state.

There are also many data sets which are mentioned as available, but which are not mentioned as needed. It may be that people don't know these data sets exist. Such lesser known, but valuable data sets might be: SRBC water use projects, mine subsidence, Act 2 land recycling sites, active/non-active landfills, emergency service zones, harvest data by county, and WCO districts.

4.1.3 Recommendations

The recommendations below address the needs defined under Parts 1 and 2, (the user and the provider) in the Analysis section above.

Recommendation 1—Focus on Inventorying Available Data Sets from State and Local Government Sources

Many data sets requested here should be contenders for a concerted effort by PASDA to build complete statewide collections, if data exists statewide, or to acquire the data for the spatial extents that are available. PASDA should place a concerted effort on investigating and documenting data sets available from state agencies and local government sources. Initially, abbreviated metadata should be created for the data and posted online to support searches, so that users are aware of the data's existence and the data's source and contact information. Later, PASDA could acquire and upload the data sets and develop more comprehensive metadata. Focusing first on an initial inventorying and documentation effort facilitates a more rapid evolution and expansion of the knowledge base for what data are available within the Commonwealth.

DESIRED SCALES AND SPATIAL ACCURACY OF SOURCES

Identifying the data scale and accuracy characteristics that are most desirable allows PASDA to focus on data sources that might best meet user needs. The questions for this section were designed to define:

1) The range of scales in which users typically work,
2) The willingness of users to use data of unknown accuracy if it is the best available,
3) The minimal horizontal accuracy necessary for user needs,
4) The minimum vertical accuracy necessary to meet user needs, and
5) Whether local data is important for PASDA to include.

4.1.4 Present PASDA Approach

Currently, data are provided by PASDA as received in terms of scale and spatial accuracy. The metadata typically informs the user about the source's scale and the data's spatial accuracy if known.

4.1.5 Analysis of the Survey Results

1) Scales Most Commonly Used:

    - For the question, "As a data user, in what range of scales (source material/output) do you typically work?" the response as tallied across the three groups is:

Scale Range	# of Responses
1:600 – 1:4800	18
1:4800 – 1:12,000	18
1:12,000 – 1:50,000	21
1:50,000 – 1:100,000	8
1:100,000 and above	5

2) Flexibility in Using Best Data Available:

   - For the question, " As a user, which statement reflects your willingness to use data that does not specifically state conformance to the National Map Accuracy Standard (NMAS)?" the responses are as follows:

Statement	# of Responses
Frequent use since nothing else is available	7
Frequent use since accuracy is not important to my applications	1
Occasional use if nothing else is available	18
Rarely use the data	4
Never use the data	2
Don't know about the NMAS	10

3) The Minimum Horizontal Accuracy Required:

   - In all cases the users chose the greatest horizontal accuracy level for their minimum accuracy requirements.

4) The Minimum Vertical Accuracy Required:

   - In all cases the users chose the greatest vertical accuracy level for their minimum accuracy requirements.

5) The High Importance Users Place on Local Data:

   - In Appendices A, B, and C, when asked the question "Is the PASDA data library complete without a comprehensive set of local government data, metadata, or links to local government Web sites?" most respondents stated emphatically that they thought it was important for PASDA to include local government data.

4.1.6 Recommendations

Recommendation 1—Establish Policy of Providing Information on All Data Providers, Regardless of the Type of Data Access

The survey results indicate that people want access to the most accurate data possible but are willing to use less accurate data when nothing else is available. The majority of respondents requested larger scale map sources similar to those created by local government.

PRIORITIZED SPATIAL PARTITION STRATEGIES (DATA TILES)

To define the best acquisition and distribution strategies, it is important for PASDA to know the partition strategy providers are currently using to store their data, and also what the preferred geographic unit is for downloading. The questions for this function were designed to define:

1) The predominant partition strategy employed in user organizations,
2) The preferred geographic unit for data downloaded from PASDA, and
3) The variations in tiling schemes that PASDA may confront when appending tiles into continuous themes within a spatial data server.

4.1.7 Present PASDA Approach

PASDA currently loads data as delivered by the data provider. If the data are partitioned into subsets such as county or watershed files, they are loaded as separate files for a given theme.

4.1.8 Analysis of the Survey Results

1) The predominant spatial partitioning strategy employed in user organizations:

Partitioning Strategy	# of Responses
County	11
Municipality	5
Watershed	5
Other	4
Not Tiled	3
USGS Quad Grid	2
Other Uniform Grid	2
Agency Districts	1

2) The preferred geographic unit for data downloaded from PASDA:

Geographic Unit	# of Responses
County	14
Watershed	11
USGS Quad	11
Municipality	8
Statewide	6
Other	2
District	1

3) The variations in tiling schemes that PASDA may confront when appending tiles into continuous themes within a spatial data server:

   - PASDA receives data in many tiling schemes now and, unless major changes are implemented, will continue to do so. Most agencies maintain their data in separate files based on some tiling scheme. For this reason it is easy for

inconsistencies to develop between tiles for a given theme. For example, a given data field for roads might be formatted differently in one tile compared with another tile. Such inconsistencies can go unnoticed if data are stored in separate files as is typically done with tiles, but can become a problem when the data must migrate into a single continuous framework.

- If PASDA chooses to implement a scalable web-based spatial data server for storing and displaying statewide data (Section 5.4), it will be critical that appending data for a given layer into a continuous theme does not become a major task for PASDA. Inconsistencies in any of the following areas could pose significant problems for PASDA in being able to quickly and semi-automatically load data into the server:
    - Attribute data structure,
    - Projection and datum,
    - Coordinate units, and
    - File format.
- For many data sets maintained as tiles, important metadata varies from tile to tile and should be captured before the data are loaded as a single layer on the web server and given a single metadata record. An example of important metadata which varies from tile to tile is the spatial extent of the tile and the date of last update for the tile.

4.1.9 Recommendations

Recommendation 1—Set Guidelines for Data Providers

While PASDA does not want to place limitations on data providers, it will need to set guidelines for data consistency. If it takes too much time to process tiled data into conformity to append the data into a continuous theme prior to loading into the spatial data server, the up-front processing will become a bottleneck to the clearinghouse's functionality. In these situations, PASDA should simply load the separate tiled data as system files on a file server, as is current practice. Users can still search the metadata and download individual tiles, but they will not be able to perform data previews or customize the download options. In no case should PASDA become involved in cleaning up the data or standardizing the attribute data structures prior to uploading data on the server, since this will detract from fulfilling the primary clearinghouse functions.

Guidelines for data providers should cover consistency issues for the following areas if it is intended for the data to be appended into a continuous spatial layer within a data server:

- Metadata—Documentation should be sufficient to allow proper data loading to the system. This will need to include the information for areas below marked critical.
- Attribute Data Structure—Critical for loading.
- Projection and Datum—Critical for loading.
- File Format—Critical for loading.

- Geographic Coordinate Units—Critical for loading.

- Resolution for Imagery—Critical for loading.

- Georeferencing Scheme for Imagery—Critical for loading.

- Number of Imagery Bands—Critical for loading.

- Edgematching—This can mean many different things depending on the type of data. Generally, this means that polygon and linear features must remain spatially continuous across tile edges and feature attributes should be consistent for the same feature that crosses a tile edge. Providers should be made aware that if they want the data to be edgematched and seamless in a continuous spatial framework, they will have to address this before providing the data to PASDA.

- Scale—Important for consistency. If scales differ, there should be a means to track this from tile to tile, perhaps at a feature-record level.

- Currency—Important for consistency. If update frequencies differ, there should be a means to track this from tile to tile, perhaps at a feature-record level.

- Horizontal Accuracy—Important for consistency. If horizontal accuracy differs, there should be a means to track this from tile to tile, perhaps at a feature-record level.

- Vertical Accuracy—Important for consistency. If vertical accuracy differs, there should be a means to track this from tile to tile, perhaps at a feature-record level.

Recommendation 2—Develop a Strategy for Maintaining Key Tile-Based Metadata Information for Data Loaded to a Single Layer on the Server

Several approaches might be adopted to track metadata differences for tiles appended into a continuous spatial data layer. One approach might involve attributing individual features with a metadata source code. Each feature record then relates to a record in a metadata table documenting the source and accuracy information. One record in the metadata table is associated with many features in the continuous spatial layer that have the same metadata source code. While this approach works with vector data, it has obvious limitations for image data sets.

An alternative approach is to create a tile-boundaries layer, then attribute the tile boundaries with the metadata associated with the tile. For example, assume a data set of streams tiled by county boundary. The metadata for each tile could be parsed into a table and linked to the county boundaries (the original tile boundary). The metadata is now associated with a graphic feature or spatial extent. Using a spatial approach to tracking metadata for original tiled data supports both continuous vector and raster data.

PRIORITIZED MAP PROJECTIONS, DATUMS, AND COORDINATE SYSTEMS

To define the best on-demand coordinate conversion services, it is important that PASDA gain insight of the projection spaces providers are presently using to store their data and the preferred projection space for customized downloading options. The questions for this section were designed to define:

1) The coordinate spaces employed in user organizations and

2) The preferred coordinate space for data downloaded from PASDA.

4.1.10 Present PASDA Approach

At the moment PASDA uploads data in the same projection/datum space as delivered by the data provider. Because data projection and datum transformation can be cumbersome for those unfamiliar with the subject, being able to provide data in the user's projection of choice will be a useful service.

4.1.11 Analysis of the Survey Results

1) The coordinate systems that are used for storing data include:

   - PA State Plane North
   - PA State Plane South
   - Polyconic
   - UTM 18
   - UTM 17
   - Lambert
   - Albers
   - Lat Long (Geographic)
   - NAD83
   - NAD27

2) The preferred coordinate systems for downloading include:

Coordinate System	# of Responses
State Plane	24
UTM	16
Other	10

3) The preferred datum for downloading include:

Datum	# of Responses
NAD 1983	35
NAD 1927	12
Other	2

4) The preferred units for downloading include:

Units	# of Responses
Feet	32
Meters	20
Other	2

4.1.12 Recommendations

If data are provided to PASDA on a consistent and well-documented basis, it should be possible for PASDA to develop automated tools to re-project data into the standard projection space required by the spatial data server. If the data are delivered in a tiled structure, it is important for the data provider to provide all tiles in the same consistent projection space. Options for automating user-defined projections for customized downloads will depend on the technology chosen to implement the system.

Recommendation 1—PASDA will Need to Choose a Coordinate System Space to Implement a Web-Based Spatial Data Server (Section 5.4)

PASDA will need to define a uniform projection space for creating the spatial framework for a web-based spatial data server, including a datum and coordinate units. While the logical choice might be geographic space (decimal degrees), it should be cautioned that many spatial-search operators require coordinates specified as ground units (feet or meters), and spurious results might be realized if units are provided as decimal degrees.

A promising projection space would be one that mimics the Pennsylvania State Plane system, as closely as possible, but does not require separate north and south zones. The Pennsylvania State Plane coordinate system is based on the Lambert Conformal Conic projection but it breaks the state into two zones (north and south). Each zone has a unique set of two standard parallels that minimize projection distortion within the zone. A Lambert Conformal Conical projection with two standard parallels equally dividing the state into thirds should be considered a candidate projection. The NAD 1983-adjusted datum would be the appropriate datum for projection, as it is the most current model of the spheroid. Units should be in meters in keeping with the move toward metric measures. However, users should be able to request spatial searches and measurements to be returned in English units.

Recommendation 2—Evaluate Technologies that Support Customized Projections

PASDA should begin reviewing and benchmarking server technology that can be used to implement on-the-fly projections of data stored in the spatial data server. Options to be evaluated include projecting the data prior to extraction from the spatial data server environment versus extracting the data as a system file and then projecting it to the user-selected projection.

DESIRED DATA FORMATS

In order to support automated format conversion, it is important for PASDA to know what formats are being used by providers and what formats users would prefer to receive data in for data downloaded from PASDA. The questions for this section were designed to define:

1) The frequency of GIS software packages and data formats that are in use and

2) The formats of choice for downloading raster and vector data.

4.1.13 Present PASDA Approach

PASDA provides data in the same format as supplied by a data provider. To date, this has typically predominantly been the ESRI export format (.e00) and shape file format for vector data, and TIFF/GeoTIFF and MrSID for raster data.

4.1.14 Analysis of the Survey Results

1) Below are user votes for type of vector and raster format for download:

Vector Data	# of Responses	Raster Data	# of Responses
ESRI Shape File Format	16	GeoTIFF/TIFF with World File	14
ESRI Coverage (.e00) Format	10	JPEG Format	6
AutoCAD (.dwg) Format	2	MrSID Format	3
Microstation (.dgn) Format	2	BIL Format	3

2) Below are the results of a survey of the GIS software organizations use to create and store raster and vector spatial data sets:

GIS Software	# of Responses
ArcView	15
Arc/Info	10
MGE	3
AutoCAD	3
GeoMedia	2
Microstation	1
Erdas	1
Ermapper	1
MrSID	1

4.1.15 Recommendations

Recommendation 1—Evaluate Technologies that Support Customized-Format Conversions

PASDA should begin reviewing and benchmarking server technology that can be used to implement binary conversions of formats for data stored in the spatial data server. Options to be evaluated include specifying a format as the data are extracted from the server or extracting the data as a defined-format system file and then reformatting the data to the format requested by the user. It may be that not all desired formats can be supported but technologies exist that provide a wide variety of binary-level format conversion options.

MAINTENANCE OF HISTORICAL VERSIONS OF DATA AND METADATA

When a new data set version is provided to PASDA the old one is typically discarded. PASDA needs to determine whether these previous versions are valuable enough to warrant maintaining them online or, at least, offline. The questions for this function were written to define:

1) Whether users would like PASDA to store previous versions of data and metadata and

2) What the best approach might be for tracking previous versions of data sets.

4.1.16 Present PASDA Approach

In most cases PASDA does not maintain previous versions of data on the site. If earlier versions exist they are kept under the Archive Data Catalog.

4.1.17 Analysis of the Survey Results

1) There was agreement in the survey results that maintaining the current and previous data versions would be desirable, but that maintaining all versions would not be necessary:

Version Maintenance Required	# of Responses
Maintain previous and current version	25
Maintain all versions indefinitely	12
Maintain current version only	7

2) The approach most favored for identifying previous versions was:

Favored Approach	# of Responses
Add version numbers to data sets similar to software versions	22
Use date of last revision in the metadata	18
Other	3

4.1.18 Recommendations

Recommendation 1—Make Previous Version(s) of a Data Set Available Online if Requested

PASDA should retain a log of all user requests for historical data versions in order to identify specific data sets that should be retained online as historical data. Only the historical versions of data sets that are specifically requested by users should be maintained online since there are significant system overhead costs of maintaining a historical and current version of every data set. To automatically keep the previous version of every data set, regardless of whether it is ever requested or used, consumes unnecessary disk space and requires creation and maintenance of two metadata files.

Recommendation 2—Focus on Creating Historical Versions of Image Data

Image data sets are most likely to be of interest for historical analysis of changes in land use and land cover over time. Image processing software can be used to evaluate change over time, and this can be used for farmland preservation studies, analysis of development trends, or loss of habitat assessments. For this reason, PASDA should focus on implementing online access to historical image data sets for the Commonwealth.

Recommendation 3—Use Metadata to Track Data File Versions

Since finding data in PASDA relies upon the metadata, there would need to be a separate metadata file created for every version of the vector or raster data. The file names for all versions of current and historical data set versions must be distinguished from one another by the addition of version numbers or dates to the file names.

5.0 OPERATIONAL IMPACTS OF RECOMMENDATIONS

Spatial data clearinghouses are new and, as a result, the approach to managing a clearinghouse is based on little experience. It is important to realize that operational issues relating to spatial data libraries have always been difficult (think about the expense and effort a map library goes through to maintain a collection of hardcopy maps) and this problem has become even more complex as a result of the intangible nature of digital data and the rapid technological changes in the GIS field. In addition, greater human and computing resources will be necessary to expand and further develop the clearinghouse services and functions as defined in the needs analysis. This Section briefly presents current operations and the likely impacts to the operations arising from recommendations in Sections 3 and 4 of this document.

PASDA SERVICES

5.1.1 Current Operations

PASDA current services are based on the clearinghouse model originally used by many clearinghouses on the National Spatial Data Infrastructure (NSDI). As the need for more data and services increases, the traditional model for clearinghouses will expand to include a variety of services and functions. The current process for creating metadata relies heavily on PASDA staff identifying and contacting data providers in the Commonwealth. Once contacted, PASDA will visit the data providers in person or begin a series of phone calls to develop full level metadata. Although PASDA provides an online metadata submission form, this tool is rarely used. PASDA metadata staff use a number of tools to create and integrate the metadata. The current approach also requires PASDA staff to convert text file metadata to XML, which is then converted, "on the fly" into HTML for display on the web site. This is accomplished by parsing the XML files through style sheets to generate a full metadata display. Metadata training is currently available from one of the PASDA staff on a first come first serve basis. Metadata training requests are forwarded to the metadata coordinator and then sessions are scheduled. The training sessions focus on the use of the online "plain English" metadata entry form available on the PASDA website. This approach is taken to encourage the creation of metadata by the data creator. Access to "as is" data is provided by ftp. Some "standard" data is also provided through the Explore PA section of the web site.

5.1.1.1 Data Cataloging (Creation of Metadata)

In order to meet the demands of the clearinghouse as defined in the analysis, efforts to train, raise awareness, and provide easy access to tools must be undertaken. For PAGIC members, tools must be identified, templates created, and education in metadata creation should be a priority of the PASDA metadata staff. By creating templates for each PAGIC member and other who submit data to the clearinghouse, time and effort will be saved. This will also allow the data providers to manage their data more easily. Data providers must also be able to update metadata when updating data.

5.1.1.2 Metadata Training

Regularly held metadata training sessions are imperative to reinforce the significance of metadata creation. These should be held in various locations across the state.

5.1.1.3 Data Inventory

A statewide data inventory should be implemented as soon as possible. This inventory should focus first on state agencies and local governments. A data survey should be available in paper and on the PASDA site for ease of use for data providers. A baseline of framework data and metadata should be available to all PASDA users.

5.1.2 Operational Impacts

The success of these efforts will rely on the participation of PAGIC members and other data providers in the state. To achieve these goals it will be necessary to have additional staff and procedures to perform the bulk of these tasks.

PROMOTION (OUTREACH)

5.1.3 Current Operations

PASDA strives to educate data users, providers, and potential users through it's mailing list, presentations at conferences, seminars, and meetings. In order to expand the user base for the clearinghouse and to ensure that all state agencies, local governments, and other data providers are aware of PASDA services, the PASDA staff must increase their outreach efforts. This increase could include working with associations, participating in government GIS user group meetings, and the development of useful resources on the PASDA site, expansion of the mailing list, and the creation of promotional materials that highlight the services of the clearinghouse.

5.1.4 Operational Impacts

The expansion of outreach efforts will benefit all data providers in the state. Working with teachers, organizations, and governments will enhance the relevance of the data and maximize the cost savings to the Commonwealth by reducing the creation of redundant information.

PARTNERSHIPS

5.1.5 Current Operations

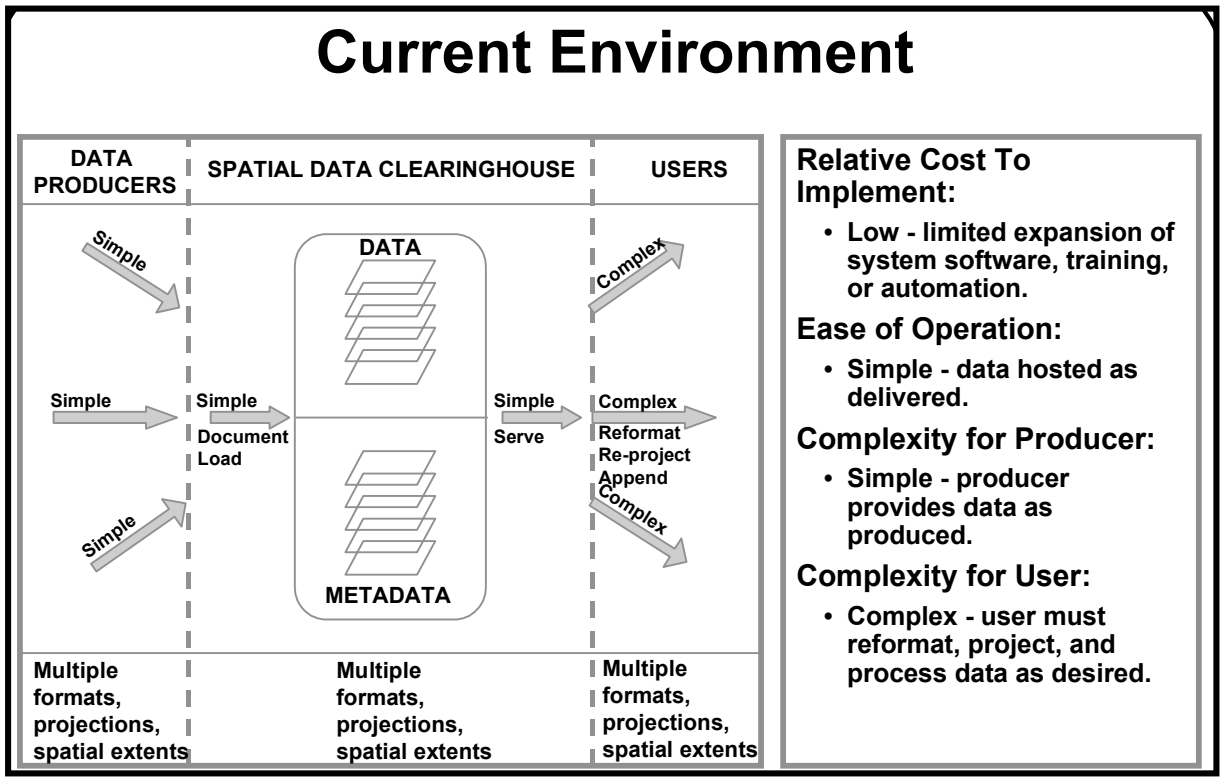
PASDA currently maintains informal partnerships with a number of organizations. These partnerships entail metadata support, data storage, and PASDA training for members. Examples of these partnerships include the Susquehanna River Basin Commission, the Delaware River Basin Commission, ALLARM, Southwest Pennsylvania Commission, SEDA-COG, and members of PAGIC. In order to more fully serve the Commonwealth and our partners, PASDA must expand it's partnerships and extend it's services to a greater number of organizations. PASDA's partnerships can facilitate inter organizational sharing of data , resources and knowledge and can assist in dealing with issues such as sensitive data, joint data development projects, and data sharing agreements. Successful partnerships with a wide range of organizations may also benefit vital functions in the Commonwealth, including emergency management operations (awareness of and access to data is vital to emergency management efforts).

5.1.6 Operational Impacts

PASDA will need to develop and implement processes and procedures for creating official data partnerships. Agreements between partners should be more concrete. The development of official data liaisons for each organization will provide PASDA with the opportunity to develop these partnerships.

5.1.7 Current Operations:

At present, PASDA accepts data in many different map projections, vendor formats, and spatial extents. The data are compressed and loaded as system files. Generally speaking, the data are loaded "as received", although PASDA has created a core set of standard data by projecting road centerlines, watersheds, streams, and floodplains into standard projections so that when downloaded, these data sets are properly aligned in a common coordinate space. The user downloads the compressed data files using File Transfer Protocol (ftp), and then must perform one or more of the following tasks:



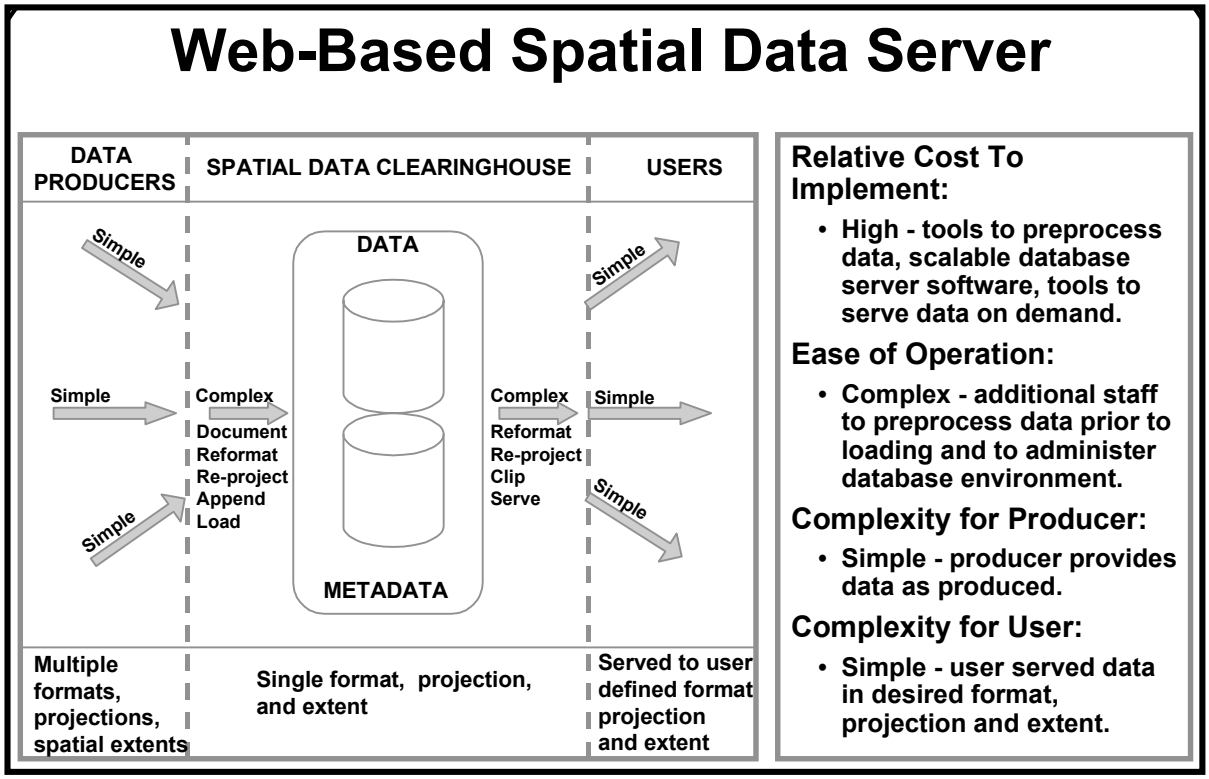
- convert the vendor format to be compatible with the user's software environment
- project the data into the correct projection space and coordinate units
- clip the data to the study area or append multiple adjacent tiles to create a "continuous" data layer.

While the current environment simplifies the clearinghouse operations, the architecture places a complex set of post-processing requirements on the user. In this architecture the clearinghouse's responsibilities are limited to data inventorying, cataloging (creating metadata), and serving "native format" data files as delivered by the data provider. There is little or no attempt to "integrate" or "standardize" the data before providing the data to the user. This means that the majority of the complexity involved in scrubbing the data is inherited by the end user. This approach is acceptable for experienced and knowledgeable GIS users (roughly 5% of the user community) who have access

to the necessary software tools and knowledge to process the data. But, it presents difficulties for novice users who do not understand conversion between map projections, reformatting of data between vendor environments, or do not have tools to perform the necessary processing.

5.1.8 Operational Impacts

Most of the individuals providing input to this study expressed a desire to be able to interact with the data on the server in order to customized the data to their needs prior to downloading the data. The user's desire is to off-load the processing to the web server, instead of being required to have the knowledge and tools to process the data after download. Fortunately, recent advances in web-based GIS and Relational Data Base Management (RDBMS) technology offers PASDA the opportunity to implement many of the requests identified by the user community in Section 3.0, though there will be significant customization and development effort. The illustration below shows the representative architecture for a new clearinghouse paradigm. This architecture is implemented using a spatially enabled object-relational data model which affords improved spatial data integration, data management, and data serving functions.



The above architecture offers several key benefits to the clearinghouse and the end user:

- Overall improved data management and administration, including data security, versioning, and data recovery from system crashes.

- An superior indexing mechanism that is better suited for spatial proximity searches accessing small areas within a large spatial extent, resulting in very fast search and retrieval times for spatial queries on very large statewide spatial data sets.

- As user demand increases, the system can be scaled-up to increase the number of data servers, thus retaining overall performance.

- Support for on-line interactive web based GIS, allowing preview of data sets and interactive mapping and analysis functions that may remove some of the need for users to download data sets to answer basic geographic questions.

5.1.8.1 Core Enabling Technologies.

Environmental Systems Research Institute's (ESRI) ArcInfo 8 contains features significant to implementing the spatially enabled data server which is the core of the system. This recently redesigned software allows users to add behavior, properties, rules, and relationships to their spatial data which is stored in a commercial RDBMS. Using Arc Spatial Database Engine (ArcSDE) as the interface to a commercial RDBMS, the ArcMap and ArcCatalog software can load, query, extract, and update spatial data within an RDBMS. ArcCatalog also has very sophisticated tools for creating and updating metadata. The metadata tools can be set to automatically update the metadata anytime a change is made to a data layer in the data server.

ArcSDE is a spatial data application server that combines the client/server architecture with a set of software services to perform fast spatial searches and extractions on a very large data sets. Key benefits include the ability to rapidly serve GIS data to a large number of users simultaneously, and simplified data management by integrating tabular and spatial data within a commercial RDBMS such as Oracle, Microsoft SQL Server, IBM DB2, Informix, and Sybase. The RDBMS plays a significant role in maintaining database security and integrity within the clearinghouse.

ArcIMS provides the clearinghouse with a Web-based GIS interface, allowing users to access published "map views" on the web server. The data for the views are accessed from ArcSDE as well as from ESRI shape file and ArcInfo coverage formats. ArcIMS enables the clearinghouse to serve map views either as "bit map" (i.e. image snapshots) or via intelligent vector streaming. As a user selects a data set for on-line previewing, the system can load the requested layer into the map viewer and classify the theme based on a predefined data item. The user can then determine whether the data set has adequate resolution and feature density for desired study area before downloading the data. Similarly, allowing the user to view the attributes of features within the theme to assess the level of attribution associated with the theme.

It can be argued that by providing a very flexible interactive mapping interface such as ArcIMS to enable users to select themes, form attribute queries, and perform basic spatial operations such as address matching and buffering, many users might not even need to download the data sets. For the general public and novice users, this functionality may well alleviate the need to purchase, install, and learn GIS software in order to answer basic geographic questions.

5.1.8.2 Tools to Standardizing Data

The data providers typically provide a variety of vendor formats, projections, and spatial extents including data tiled by county or watershed. The proposed architecture requires the data to be loaded in a consistent projection (see Section 4.4.3), defined in a common data format such as an ESRI shape file or ArcInfo coverage, and appended into continuous layers to remove embedded tile boundaries. This presents some obvious concerns regarding the level of automation that can be achieved on the processing, since "standardizing" the data could easily become a bottleneck. Fortunately, a variety of commercially available GIS software packages and third party tools exist that could be adapted and used to automate many of the preprocessing steps. The selection of the proper tools should be based on benchmark testing, perhaps as part of a prototype:

- Reformat vendor formats - Arc 8.0 (ESRI) and GeoMedia 4.0 (Intergraph) both have tools for reading and writing different formats. Also, FME Professional (by Safe Software Inc.) supports 43 formats of vector and raster data. Ideally, the selected tool should be capable of direct reading and writing binary files of various formats without the need to use intermediate ASCII formats.

- Project - Arc 8.0 (ESRI) supports conversion between many different map projections and datums.

- Append - Arc 8.0 (ESRI) can append multiple files into continuous tile and mosaic a variety of image and vector formats.

- Create new metadata for seamless data layers from multiple source metadata files relating to original tiled data. Custom programming is likely to support this function.

5.1.8.3 On-demand Data Customization

The users generally like the idea of having the flexibility to customize the data on the server side to meet their specific needs (Sections 3.3 and Section 4). Some of the geo-processing tools required to extract and clip data to a subset area, re-project to a user specified projection, and reformat the data are available within web-based GIS packages, while other web enabled tools can be purchased from third party vendors. Regardless of which tools are implemented, it will require considerable customization of both the client browser view to add in the desired function calls, and development of server side programs that get executed in response to the request. The selection of the proper tools should be based on benchmark testing:

- Clip - Possible use of ArcIMS with browser customization to interface with the ExtractServer

- Projection - Possible use of ArcIMS, but requires additional options added to the ExtractServer.

- Reformat - Possible use of ArcIMS, but requires additional options added to the ExtractServer. Alternatively, FME Objects (by Safe Software Inc.) supports 43 formats, connects directly to SDE, and is supported by many Object Component Models (C, C+, COM, Java, etc.)

- On-the-fly compression -Winzip32 or equivalent compression software

A related issue resulting from allowing user's to customize the data stored on the clearinghouse site is that custom tools must be created to modify the metadata "on-the-fly" to reflect the spatial extent, projection, and format information requested by the user. Otherwise, the metadata files that are downloaded will not properly reflect the data set that is downloaded.

MANAGEMENT

Clearinghouse management is defined as the practice of establishing policies and procedures, implementing improved technologies, and setting up work schedules for controlling the data operations. Five dimensions of clearinghouse operations were evaluated to determine the impact on management and staffing. These are:

- Scope—Scope ranges from a state agency data file to a statewide relational database where spatial data are shared among business functions. Currently the scope of the PASDA Clearinghouse is a centralized repository of state agency data files and it is moderately complex to manage. The management, technical, and physical infrastructure demands will become much more complex if a shared relation database architecture is assumed. In addition, the increase in responsibilities to a wider user base, increases the complexity and frequency of tasks performed by the clearinghouse staff and management.

- Spatial Extent—Spatial extent refers to the size or area the data covers. The clearinghouse should work primarily with organizations within the Commonwealth but must also work with regional groups, such as the Susquehanna River Basin Commission, Chesapeake Bay Program, and others to address such issues as watershed management and land use data of significance to the Commonwealth (watershed and land use data tend to be data that cross state lines). In addition, PASDA will assist the Commonwealth in its interactions with other state geospatial data clearinghouses.

- Spatial Detail and Data Mass—Spatial detail concerns how much detail is being recorded. This is closely related to the source material's scale. Data mass refers to the volume of data covering a particular area. It is anticipated that the data mass will significantly increase as new and legacy data sets are acquired and created.

- Technology Life Cycle—The technology lifecycle is comprised of several processes—plan, define requirements, design, develop, operate, maintain, and evolve. It identifies the phase of the overall system and data architecture. Based on the rapidly changing technological environment, the lifecycle for the PASDA Clearinghouse can be anticipated to undergo a substantial turnover every two years. The technology lifecycle will have a continuous and significant impact on operational issues.

- Clearinghouse Maturity—Clearinghouse maturity indicates where the actual capability is in its development level. The successful clearinghouse will likely have policies,

directives, standards, success measures, and budgets for overall activities. Currently the PASDA Clearinghouse is in its initial development phase. The Clearinghouse is working towards a stage where processes are acknowledged that repeat initial successes. This will eventually result in defined database activities that are documented, standardized, and integrated into a standard process for the organization. These defined databases will be managed with detailed quality standards and control processes.

Though the clearinghouse management structure was not studied as part of the user input sessions, the impact of those sessions will directly impact the management and staffing of the clearinghouse including the number and type of skills of staff members. PASDA performs a number of functions related to it's role as official clearinghouse. The scope of these functions are engineered through consultation with the PAGIC and the PADEP, the clearinghouse funding agency. The Services Committee of the PAGIC works directly with the PASDA Clearinghouse Coordinator/State GIS Librarian to develop strategies, implement new services, and promote the use of the clearinghouse. All PASDA staff are considered staff of and consultants for the PAGIC. Given this, there are nine classes of PASDA Clearinghouse staff that are anticipated to be necessary based on the dimensions of PASDA Clearinghouse operations. These are:

- ClearinghouseCoordinator/State GIS Librarian—The Clearinghouse Coordinator/State GIS Librarian develops the vision, manages the staff and maintains a focus of how the PASDA Clearinghouse can efficiently accomplish goals and objectives. The Clearinghouse Coordinator works directly with the PAGIC , it's member agencies, and the Services Committee to set these goals and objectives. .

- Metadata Coordinator/Metadata Staff—The Metadata Coordinator, under the guidance of the Clearinghouse Coordinator, develops and manages the Clearinghouse metadata, supervises the staff functions related to metadata, trains users in the creation of metadata, and develop and implement metdata database management tools.

- DatabaseManager(s)—The clearinghouse database manager is responsible for ensuring that the database retains its integrity and is accountable if the database becomes compromised. The database manager(s) is responsible for updating and adding new data, checking data accuracy/curreny, and for maintaining the fulfillment of data agreements. The manager(s) interacts with the end-users to identify, document, and maintain databases. The manager(s)are the ones that recognize data inaccuracies, inconsistencies, and duplication. As such, each data set should be assigned to a staff member who is responsible for defining the element's access and security rules. Assigning ownership of specific data elements to a database manager encourages quality data maintenance, thereby enhancing overall clearinghouse integrity.

- Technology Coordinator/Specialists—Since the technology life cycle of the clearinghouse changes rapidly, it is important to maintain a full understanding of changes in technology and to integrate these changes into the clearinghouse. An individual or team of individuals who provide the coordinator and Services Committee of PAGIC with information on direction and costs related to new technologies will serve to keep the clearinghouse performing along industry lines.

- Education/Outreach/Marketing—Opportunities to expand clearinghouse services to new or existing data providers and users must be explored continuously to optimize the funding available to the clearinghouse.

- Data Liaisons—While most PASDA staff will perform a variety of tasks, it will be important to form direct and lasting connections with data providers. In order to develop these partnerships, PASDA must develop a core of data liaisons who work directly with data providers and potential data providers. These liaisons will specialize in the unique information their assigned provider develops. In addition, specialized liaisons for local and regional governments, as well as non-profit organizations such as universities and watershed groups, will be necessary to take advantage of the vast stores of data they create and to provide these groups with the assistance necessary to access and use PASDA.

- Applications and Systems Analyst (s)—Analysts access the database to set up processes. Analysts have the freedom to perform their tasks but should be constrained from altering data or system utility programs.

- System Administrator—Individual or group who maintain the "physical" clearinghouse systems and functions such as indexing software, scripts, software and hardware.

- Webmaster—Individual or group responsible for web site design, content update, and web site maintenance.

Oversight

Since the PASDA Clearinghouse's lifecycle can be anticipated to undergo a substantial turnover every two years, an annual assessment is suggested to identify strengths and weaknesses and establish goals for improving operations. The Services Committee of the PAGIC is tasked with holding the annual evaluation and, in consultation with the PASDA coordinator, implementing resulting recommendations. The assessment might include members both internal and external to the Council. Those internal to the organization should be practitioners in some aspect of GIS. The assessment could produce recommendations and serve as the basis for future goals and operational directions.

Impact

The impact of this needs analysis will have a direct affect on the clearinghouse management and staff. The expectations of PASDA's users have greatly increased and will continue to do so based on the current trends in geospatial data use and clearinghouse development. More staff will be required to meet both the technological needs (database development, design, administration, expansion, and innovations) and the outreach (education/training, data liaisons, data acquisition, metadata development, partnership development, marketing) services. PASDA management will need to define, on an annual basis, it's goals and plans for the coming year. As the clearinghouse grows in complexity and as expectations rise, the staffing and technological requirements for the clearinghouse need to grow as well.