

Pennsylvania Spatial Data Access Clearinghouse Benchmarking Summary

11/30/2014

Maurie Caitlin Kelly

Updated 12/15



**PENNSYLVANIA
SPATIAL DATA ACCESS**

The Pennsylvania Geospatial Data Clearinghouse

The Pennsylvania Geospatial Data Clearinghouse

Executive Summary

Pennsylvania Spatial Data Access (PASDA—<http://www.pasda.psu.edu>) is the official public access geospatial information clearinghouse for the Commonwealth of Pennsylvania and has served for 20 years as Pennsylvania's portal to GIS data and as our node on the National Spatial Data Infrastructure and NSGIC GIS Inventory. PASDA was developed by the Pennsylvania State University as a service to the citizens, governments, and businesses of the Commonwealth. PASDA is a cooperative project of the Governor's Office of Administration, Office for Information Technology and the Penn State Institutes of Energy and the Environment of the Pennsylvania State University. Funding and support is provided by the Pennsylvania Office for Information Technology. In addition, PASDA also receives substantial support from the Pennsylvania State University.

PASDA stores more than 80 Terabytes of data and provides access to more than 400,000 pieces of data. Note: This number includes zipped data, KMLs, and geodatabases (geodatabases are more than one data set combined into a larger database. Individual data sets within a geodatabase were not counted). The Statewide Imagery Cache (this data is only available for viewing via a map service or via the PA Imagery Navigator) is an additional 1.2 million jpegs.

- 2015 Statistics:
 - PASDA Hits: 68,517, 275
 - Page Views: 35,099,555
 - PASDA Unique Visitors: 1,041,886
 - Datasets Downloaded: 1,912,746
 - Total Maps Requested (Map Services, KMLs, WMS): 44,000,568

Benchmarking

Approximately every five years the PASDA staff undertakes a benchmarking effort to research current trends and advances in GIS data clearinghouse development. The previous benchmarking and user input session was held in October of 2008. In September of 2014 the PASDA staff embarked on such an effort. The overall goal was to identify best practices in the field as well as examine and evaluate alternatives and enhancements for the PASDA clearinghouse.

For the purposes of the benchmarking process, we researched each US state to identify a) whether or not there was an active clearinghouse and b) how data was searched and retrieved. In addition we identified several specific criteria to examine. Each staff member was assigned a particular area to research and report on. In addition, we noted the status of various state clearinghouses in the US including date of launch and longevity and quality of the website and materials available via the site

The four areas of focus for the effort were:

1. Data search and retrieval functionality
2. Data format, access options, open data
3. Applications and visualization tools
4. Preservation & Persistence

In addition, we also compared available data, imagery, and LIDAR. In October of 2014 the PASDA staff held a meeting to discuss and analyze the results. This report provides a summary of our findings related to these areas. In addition, the report was updated in December 2015 to reflect recent changes in data access architectures and to support presentations that PASDA staff prepared for the National Academy of Sciences, the Pennsylvania Society of Land Surveyors, and multiple GIS Day events.

US Clearinghouse Overview

There is a wide variety of clearinghouses in the US. Many of these clearinghouses have been in existence since the mid to late 1990s like PASDA (1996) and New York State (1995). The first efforts related to clearinghouse development and data sharing were spearheaded by the Federal Geographic Data Committee (FGDC). PASDA was connected to the NSDI and searchable through the Z39.50 protocol and Isite software - an open source program acquired from FGDC. Within three years, Pennsylvania was one of about 25 data clearinghouses to be searchable through the NSDI - others included New York, Texas and Arkansas. However, many of the clearinghouses that began during this time have gone offline or were offline for long periods of time (Alabama, Georgia). Others have changed dramatically and have potentially set up competing sites that can be at times frustrating for users looking for coordinated data portals (Arkansas, Delaware).

Among these there are multiple types of clearinghouses. PASDA has always been an open data portal which is accessible for free to any user with an Internet connection—no login or account required. New York, until last year, had always been a member only cooperative. Only members could obtain data or users could also request data through a cooperative representative (New York model; Georgia has also implemented logins and accounts though they are not limited to members). A newcomer at the state level to the coordinated GIS portal movement is the state of Maryland which launched its GIS data portal in the last two years after an extensive data inventory effort. NOTE: There were several unofficial access points for Maryland data that served many of the users in the GIS community before the Maryland portal came online.

Several states do not have a GIS data clearinghouse. For example, Alabama, which in previous years had very limited access to GIS data was offline during our initial survey. They do have an initiative called “Virtual Alabama” that provides access to 3D imagery but users must have a user ID and password account to access the site. It also demonstrates what is an ongoing issue for users—information silos.

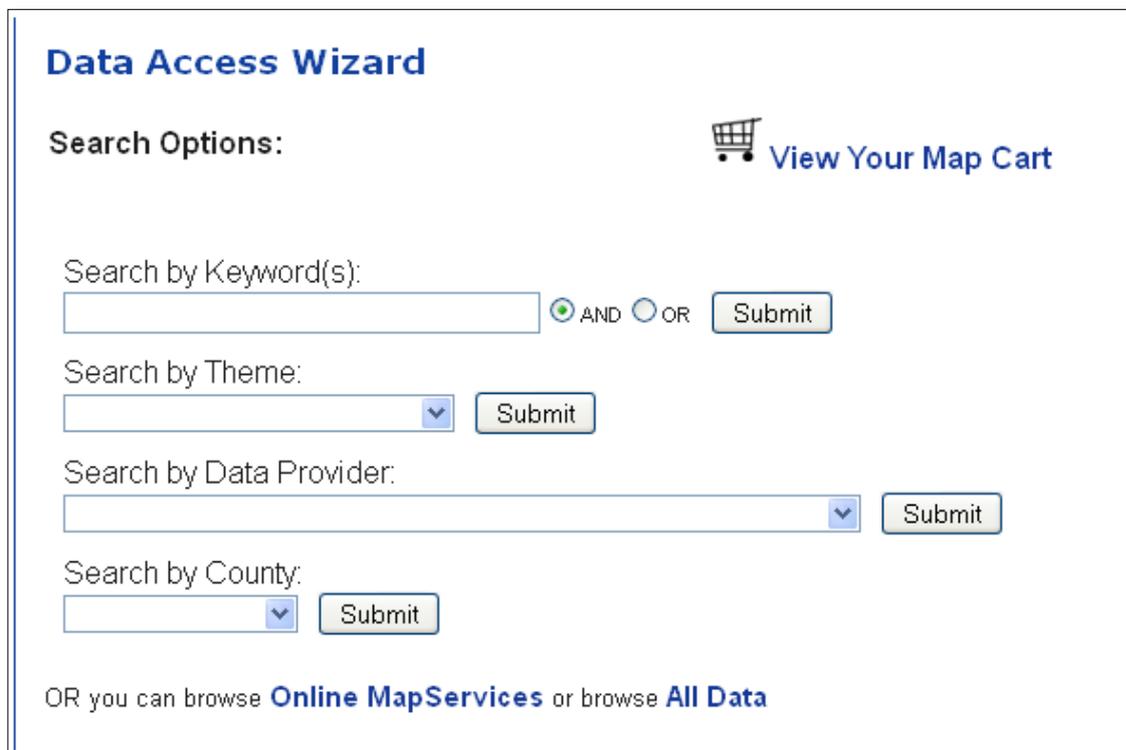
An **information silo** in this context is a state that requires users to jump from agency to agency or application to application to identify and access data. This uncoordinated siloed approach is typified

in the state of Florida where users do have access to some data but have to search multiple sites separately including: Florida Department of Environmental Protection: GeoData Directory; Florida Department of Environmental Protection: Land Boundary Information System (LABINS); Florida Department of Transportation Geographic Information System (GIS); Florida Geographic Data Library (FGDL). This process can be very time consuming and does not facilitate cooperation, collaboration, nor a positive user experience. We do not believe that this approach is conducive to open access to data which has been a strong tradition in Pennsylvania.

One of the most common comments that PASDA staff have received over the years is that PASDA is the best clearinghouse in the country. Users who have experienced trying to find data for other states find themselves often frustrated and spending a lot of time jumping from one agency website to the other and often not finding the data they need. Joseph Kerski, in his most recent book The GIS Guide to Public Domain Data, calls PASDA “the model for what a truly user-friendly and useful data portal SHOULD be”.

Data Search and Retrieval Functionality

The current data search and retrieval functionality on PASDA was developed in 2006. This interface, called the Data Access Wizard (**Figure 1**), introduced enhanced ability to identify data on PASDA. The driving force behind this effort was the dramatic increase in data over the previous four years. PASDA had acquired a great deal of new data as well as numerous data partners and it became imperative that we develop a more effective way for users to find data.



The screenshot shows the 'Data Access Wizard' interface. At the top left is the title 'Data Access Wizard'. Below it is the 'Search Options:' section. On the right side of this section is a shopping cart icon and the text 'View Your Map Cart'. The search options include: 'Search by Keyword(s):' with a text input field, radio buttons for 'AND' (selected) and 'OR', and a 'Submit' button; 'Search by Theme:' with a dropdown menu and a 'Submit' button; 'Search by Data Provider:' with a dropdown menu and a 'Submit' button; and 'Search by County:' with a dropdown menu and a 'Submit' button. At the bottom, there is a link 'OR you can browse Online MapServices or browse All Data'.

FIGURE 1. THE PASDA DATA ACCESS WIZARD

The PASDA Search Results page provides users with access to the following (Figure 2):

- Metadata summary
- Direct Download
- Access to Map Services , Google Earth KMLs, and the Data Previewer
- Access to Applications
- Map Making capabilities

Title	Originator	Date
Abandoned Mine Land Inventory Points	Pennsylvania Department of Environmental Protection	2014
Abandoned Mine Land Inventory Polygons	Pennsylvania Department of Environmental Protection	2014
Abandoned Mine Land Inventory Sites	Pennsylvania Department of Environmental Protection	2014
Active Underground Permit Boundaries	Pennsylvania Department of Environmental Protection	2014
Air Emission Plants	Pennsylvania Department of Environmental Protection	2014
Ambient and fixed station network groundwater monitoring point data (1985 - 1998)	Pennsylvania Department of Environmental Protection	1999

FIGURE 2. PASDA SEARCH RESULTS PAGE.

Other Clearinghouse Best Practice

OVERALL APPEARANCE & EASE OF USE OF SEARCH AND RETRIEVAL FUNCTION

The most user friendly site we identified in this category was the **West Virginia GIS Technical Center GIS Clearinghouse**--<http://wvgis.wvu.edu/index.php> (Figure 3). We selected this site due to the clarity of presentation, ease of access, and variety of data. In addition, this site provided more information than simply data. It also served as an access point to information about other sites, standards, applications, and GIS in WV in general. Another positive, though more subjective, aspect

of this clearinghouse is the responsiveness of the staff. They have a reputation for excellent customer service and responsiveness.

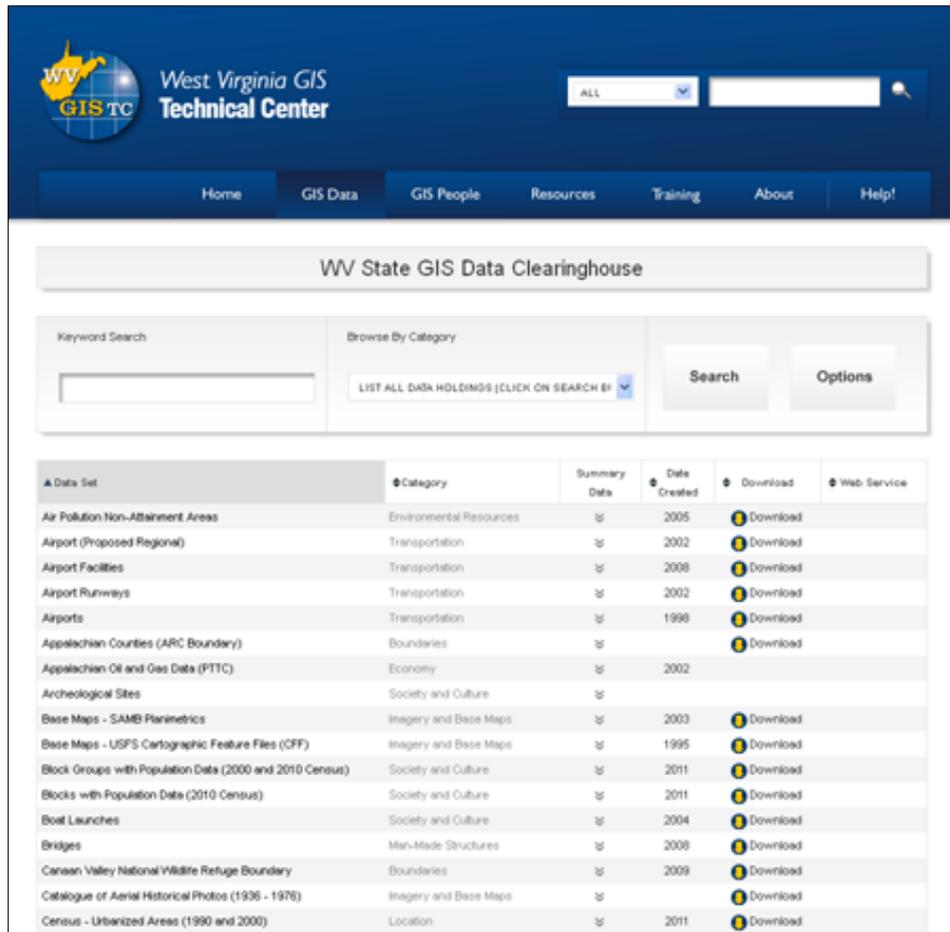


FIGURE 3. WEST VIRGINIA GIS CLEARINGHOUSE

After our review of existing clearinghouse sites and user feedback, as well as search engine interfaces such as Google, the PASDA staff determined that the most significant change that could be made would be to streamline the existing search interface and eliminate content from the home page. The goal was to get users to data more quickly and move content to other pages such as About and Events and News.

With this in mind, we developed the new PASDA homepage and search interface (**Figure 4**). The new home page serves as the first search page by providing a search by keyword, provider, and the ability to quickly browse the FTP (long time users of PASDA find this option particularly important). In addition, the new home page provides immediate access (Data Shortcuts) to heavily used data such as imagery, LIDAR, hydrology, and transportation data.

Advanced search options, such as search by Map, are still available through the Advanced Search page. Finally, we also included a section on New Additions to highlight new data and maintained our Spotlight on GIS to feature other organizations applications.

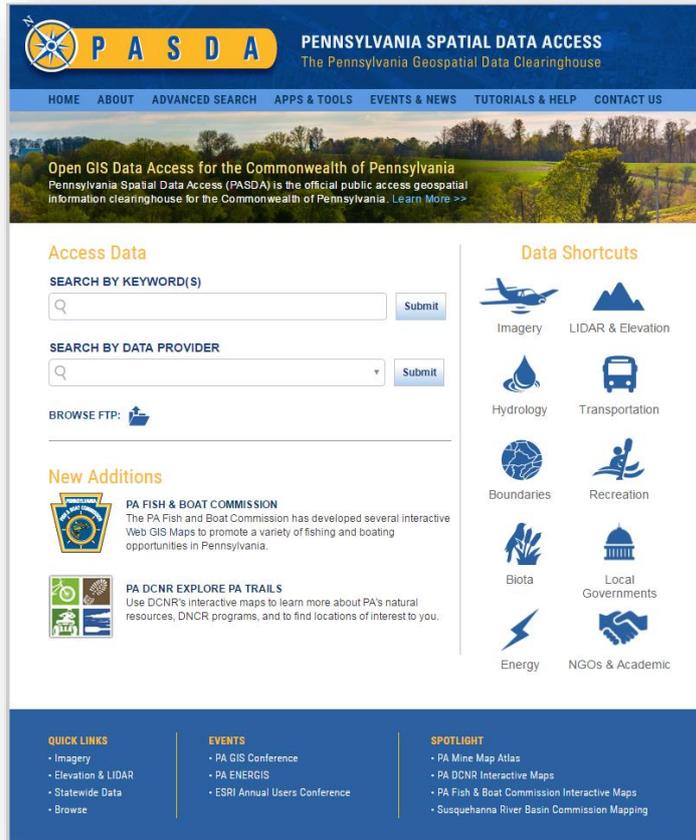


Figure 4. New PASDA Home Page & Search Interface

The second component of this effort was to address the Search Results page. This page was streamlined (Figure 5).

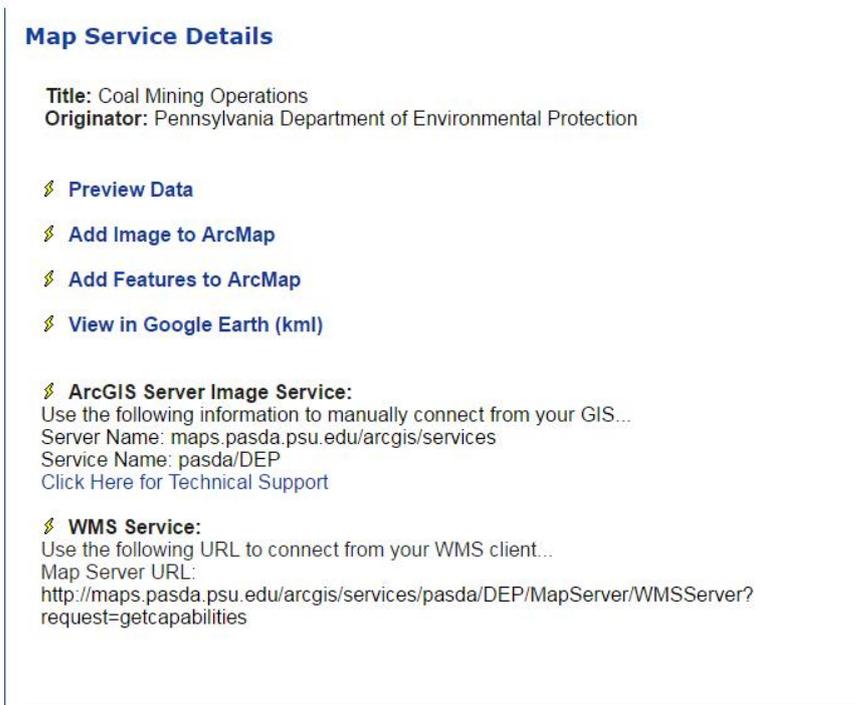
Date	Title	Provider
2016	PennDOT - Bridges	Pennsylvania Department of Transportation
2010	Glacial Deposits of Northwestern Pennsylvania	Allegheny College
2010	Thickness of Glacial Deposits, Western Crawford County, Pennsylvania	Allegheny College
2010	Thickness of Unconsolidated Deposits of Erie County, Pennsylvania	Allegheny College
2016	Allegheny County - Address Points	Allegheny County
2002	Allegheny County - Basins	Allegheny County
2002	Allegheny County - Blocks	Allegheny County
2016	Allegheny County - Boundary	Allegheny County
2016	Allegheny County - Building Footprints	Allegheny County

Figure 5. New Search Results Page

DATA FORMAT & ACCESS OPTIONS

In 2006 PASDA recognized the emerging need for providing access to Internet Map Services. These services allow users to “pull” data from PASDA into their GIS software without actually having to download and store the data on their local server. This led to the creation of the Online Map Services that include the Preview Data option, Add to ArcMap, View in Google Earth (or download KML), Image Service, Feature Service and WMS Service.

PASDA was the first GIS clearinghouse in the US to provide multiple types of access to data—in particular Feature, Image, and WMS services in addition to direct download of data (**Figure 3**).



Map Service Details

Title: Coal Mining Operations
Originator: Pennsylvania Department of Environmental Protection

- 🔗 [Preview Data](#)
- 🔗 [Add Image to ArcMap](#)
- 🔗 [Add Features to ArcMap](#)
- 🔗 [View in Google Earth \(kml\)](#)

🔗 **ArcGIS Server Image Service:**
Use the following information to manually connect from your GIS...
Server Name: maps.pasda.psu.edu/arcgis/services
Service Name: pasda/DEP
[Click Here for Technical Support](#)

🔗 **WMS Service:**
Use the following URL to connect from your WMS client...
Map Server URL:
<http://maps.pasda.psu.edu/arcgis/services/pasda/DEP/MapServer/WMServer?request=getcapabilities>

FIGURE 6. MAP SERVICE DETAILS

Other Clearinghouse Best Practice

An example of a clearinghouse that has followed this model is the Arkansas Spatial Data Infrastructure, ASDI—formerly known as Arkansas GeoStor. GeoStor (ASDI) was one of the first clearinghouses in the country, along with PASDA, New York, and Texas. In the last year, ASDI has undergone many enhancements and improvements (**Figure 7**).



Figure 7. ASDI Data Format and Access Options

However, advances in software and technology have allowed clearinghouses to develop and deploy new ways in which users can interact with data. The PASDA staff examined several newer data portals/clearinghouses to identify best practices. We identified one key area for improvement—access to additional data formats (**Figure 8**). The new PASDA site includes existing PASDA access options for direct data downloads, metadata, Preview, Feature, Image, and WMS services, and linkage to any related apps or tools, it also provides access to JSON, and CSV (where applicable).

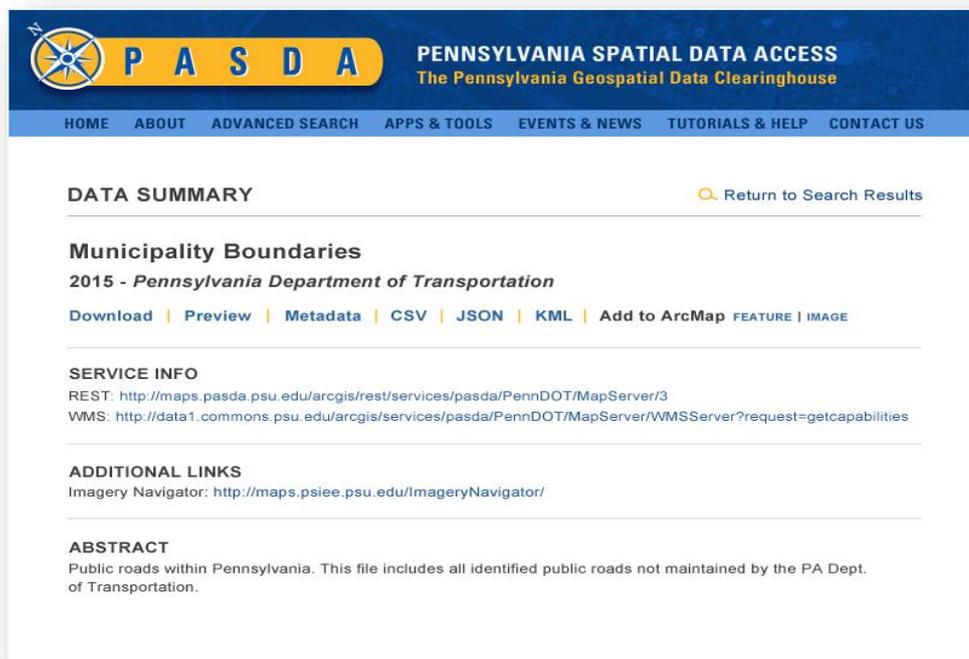


Figure 8. New PASDA Data Access Options.

A popular trend in data is the Open Data Portal. There are multiple definitions for what open data means—open and free to the public, open format, live data, etc. Cities such as New York, Philadelphia, and Chicago have developed Open Data Portals with varied success. The Open Data in these portals is any data that is generated by government entities, in particular tabular data that has often been the subject of great citizen interest but little access. For example, many government agencies make their data available as tables in reports in PDF files. Citizens may actually just want the database behind the table to analyze themselves. New York City’s data portal, which utilizes Socrata, provides access to such data (Figure 9).

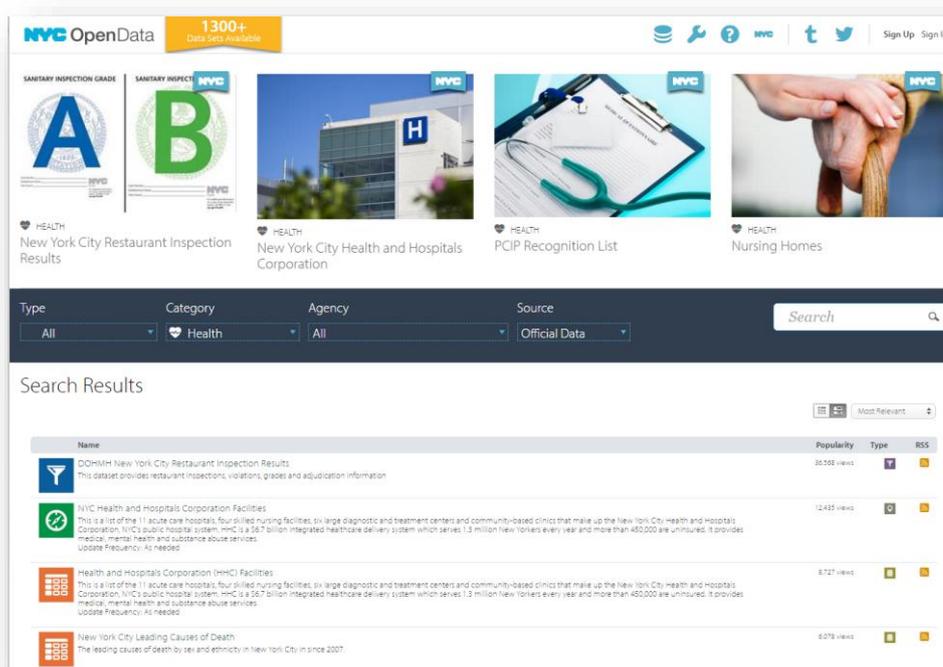


Figure 9. NYC Open Data

In addition, to providing access to tabular data the NYC site also has the option to view some GIS data in a previewer type interface and download the data. Some sites take data from organizations outside of government. Philly Data, which runs on the CKAN platform, has jobs from Craig’s List and Drink Philly Happy Hour Specials.

Data portals that specialize in GIS data are also appearing, mostly due to the development of ESRI’s ArcGIS Online Platform. ArcGIS Online is the outcome of previous efforts by ESRI including the Geography Network which was launched in 2000 and of which PASDA, along with Texas and National Geographic, was one of the first participating members; the subsequent Metadata Explorer of which PASDA also was an early adopter; and ArcGIS.com. Maryland, which is new to the open data and GIS clearinghouse movement as a state, has developed its open data portal using Socrata. PASDA has an ArcGIS Online instance and has extensively tested both our own instance’s performance and those of other AGOL instances including Maryland. Our research has identified several issues related to AGOL sites: incomplete data downloads, inability to download large datasets, time outs and errors in processing, changes to metadata, and scrambled attributes. Mark Headd, former Chief Data Officer at the City of Philadelphia and primary driver of their Open Data

initiative has many concerns about money, time, and impact and states that Open Data Portals should “Get out of my way and give me my data!”. This statement emphasizes the need for focusing on the data first—ensure its integrity, consistency, and easy accessibility first, the “bells and whistles” are less important.

Applications and Visualization Tools

In 2006 PASDA staff recognized that Google Earth, which was originally known as EarthViewer 3D, was becoming increasingly popular. While Google Earth allowed users to see data that Google hosted such as satellite imagery but there was limited interaction with locally held data. The ability to create and share KML files enabled us to integrate data on PASDA with the Google Earth interface.

In addition, PASDA supports several important applications including the Pennsylvania Imagery Navigator, the Pennsylvania Atlas, and the Pennsylvania Mine Map Atlas.

The Imagery Navigator (**Figure 10**) was developed to help users access the multiple types and years of imagery and elevation data available through PASDA.

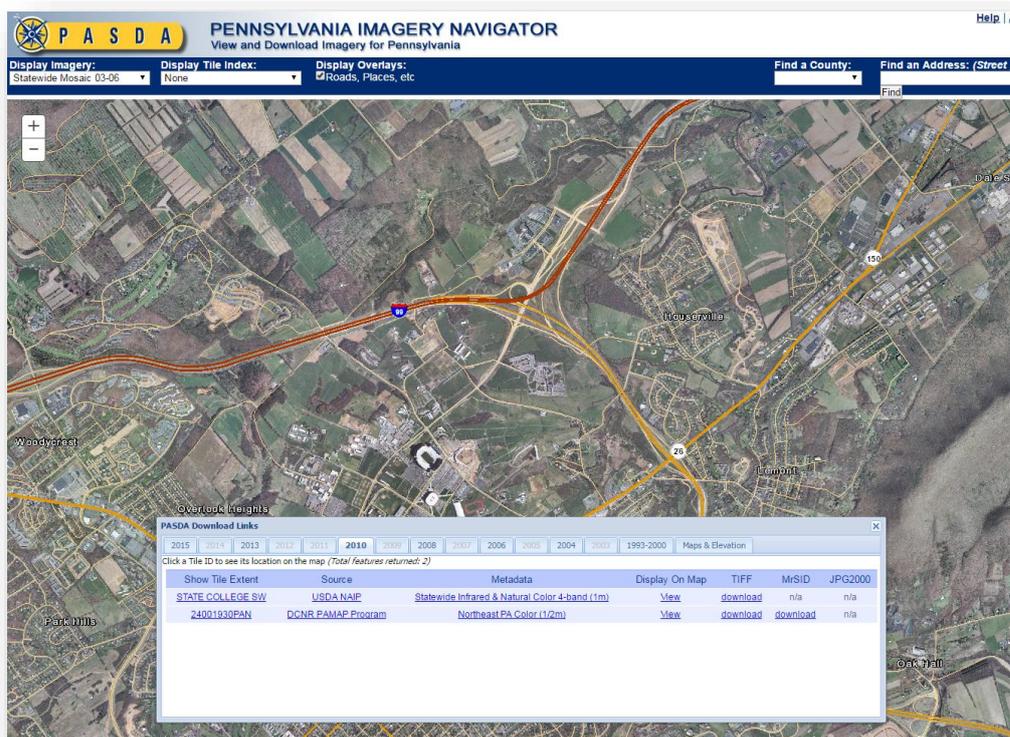


FIGURE 10. THE PENNSYLVANIA IMAGERY NAVIGATOR

The imagery available includes PAMAP, Delaware Valley Regional Planning Commission, Lancaster County, Philadelphia, Allegheny County, Lehigh Valley Regional Planning Commission, USGS, NGA, NAIP, and more. The elevation and topographic data available through the Imagery Navigator includes PAMAP LIDAR, Philadelphia elevation and LIDAR, Allegheny County elevation and LIDAR, USGS topographic maps (DRGS), USGS digital elevation models (DEMS), and more.

Other Clearinghouse Best Practice

A comparable site is the Oregon Imagery Explorer, <http://imagery.oregonexplorer.info/> (**Figure 11**). Much like PASDA, Oregon provides access to aerial photography through image services, meaning you may stream high resolution imagery online or to a GIS client without downloading large amounts of data. Users can preview any of the imagery datasets and by accessing web services—REST and WMS.

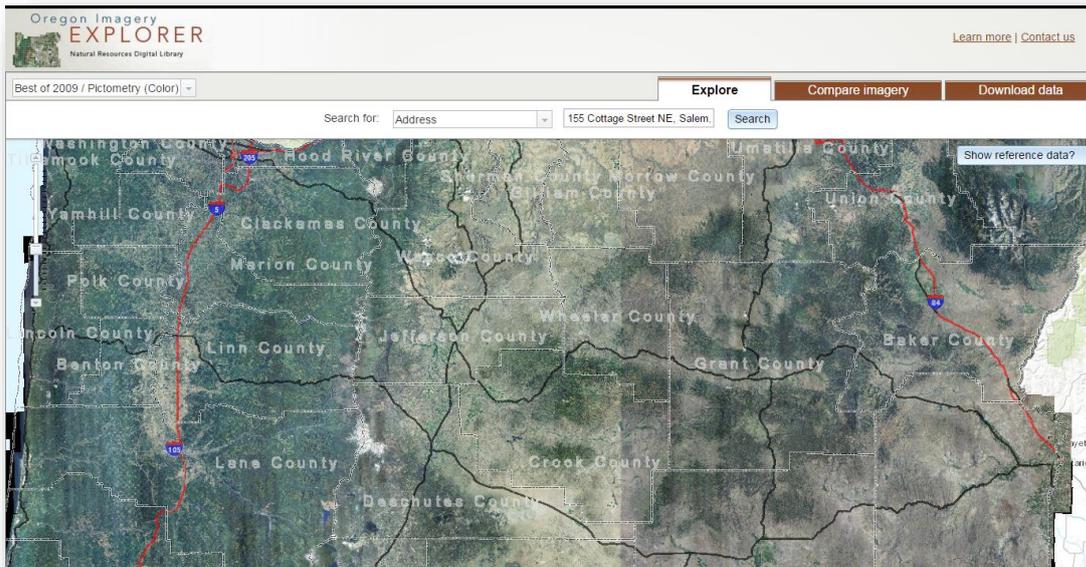


Figure 11. Oregon Imagery Explorer.

In addition to the Imagery Navigator, PASDA also supports the Pennsylvania Mine Map Atlas. This project is a cooperative effort of the Pennsylvania Department of Environmental Protection (DEP) and PASDA and is funded by the DEP. PASDA worked with the California PA Mining Office to identify needs and develop the user interface. This Web-based mapping application and downloadable data allows users to see detailed underground mine maps that were once only available in hard copies. It not only allows homeowners to view previously unavailable mine maps, but also allows them to see their home's proximity to the nearest underground mine (**Figure 12**).

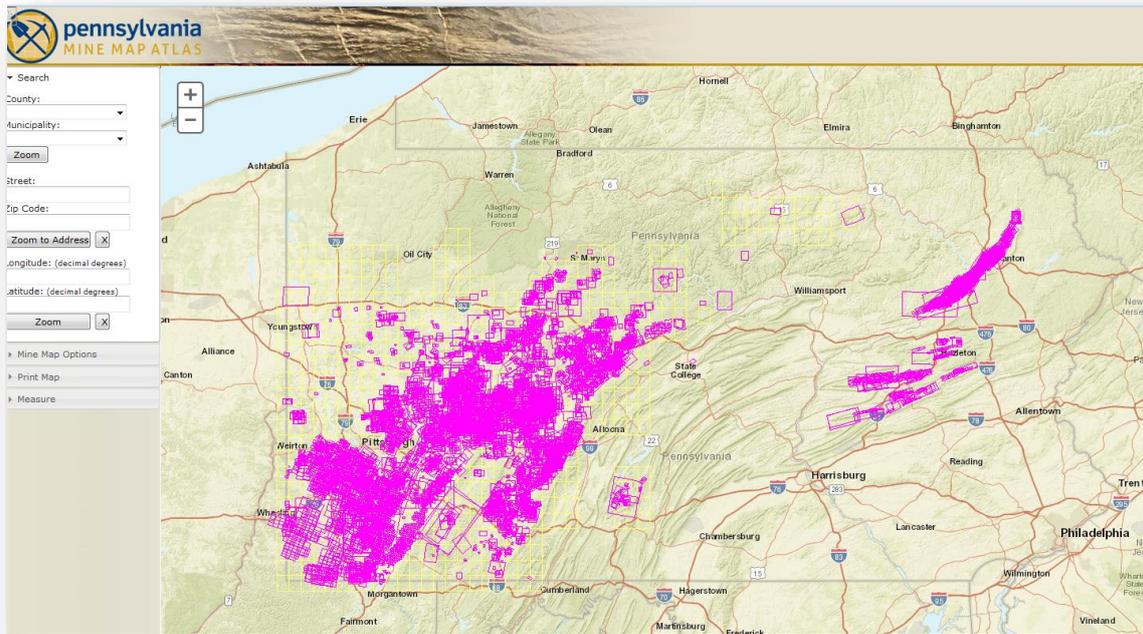


Figure 12. PA Mine Map Atlas.

Preservation & Persistence

Two issues that has received significantly less attention yet are highly significant are preservation of data and persistence of location. The Library of Congress and Federal Geographic Data Committee have been working on this issue for several years through a committee dedicated to identifying and reporting out on data preservation. PASDA is a member of this committee and contributed to the report—specifically on the section of the report on data of high scientific and research value. PASDA is committed to long term preservation and access to historic data sets for Pennsylvania. PASDA currently has a two level approach. First, PASDA maintains an “Historic” folder on the FTP site for each agency and dataset that is updated. As new data comes in, the previous data is moved to the historic folder so it can be accessed by users. In addition, in the new PASDA site, the option to access historic data will be a link in the Data Summary page. The second level is long term archiving of data. PASDA, in collaboration with the Penn State University Library and AP Trust, is working with the BagIt system which is a long term (10, 20, 30+) year archival program. All PASDA data will be placed in “Bags” and placed in near term storage. These data will be regularly scanned and checked for integrity and completeness.

The final and equally important issue is persistence. How can we maintain a persistent URL for data so users can always find it, cite it, reference it in reports and scientific works. PASDA is working with the Penn State Library to provide Digital Object Identifiers (DOIs) for each data set that will be persistent across any changes that might take place within the PASDA architecture.

We were unable to identify any other clearinghouse that has implemented an archival or DOI approach for data.