PA GIS DATA SHARING

2022 Report

Philadelphia, PA, Overlooking Benjamin Franklin Parkway

Introduction

Every day thousands of Pennsylvanians start their day by logging onto their computers, laptops, and mobile devices and accessing any number of thousands of freely available, publicly accessible geospatial data sets. Thanks to this, Pennsylvanians are able to quickly and easily do their work, run their businesses, build roads, plan communities, respond to emergencies, continue their research, conduct government business such as providing benefits and collecting tax revenues, complete their student projects, serve their patients, and help to conserve our environment. Pennsylvania was one of the first states to openly share data, and that tradition has grown and expanded exponentially over the last nearly three decades.

Overview

Pennsylvania has a long history of openly sharing data dating back to 1995 when the Pennsylvania Department of Environmental Protection (PA DEP) approached Penn State University about hosting and providing access to PA DEP data on the Internet. The agency was fielding regular requests for its data and these requests were taking considerable staff time and energy.

The timing could not have been more perfect. Just two years earlier, the National Spatial Data Infrastructure (NSDI) had been envisioned as a national level metadata and data network that provided information about data from all levels of government (3). This coincided with what was termed the "geospatial data

clearinghouse" movement, which grew out of the need to share spatial data in the early 1990s (1). These initiatives had grown to such an extent that by 1995, the Federal Geographic Data Committee (FGDC) had deployed the FGDC Metadata Clearinghouse which consisted of a gateway for participants to share their metadata, and through this, their data holdings from a single point of access (1). Pennsylvania, through Pennsylvania Spatial Data Access (PASDA), was one of the first states to become a partner in the FGDC Metadata Clearinghouse, sharing out the PA DEP data sets along with United States Geological Survey (USGS) topographic digital raster graphics (DRGs).

From the initial 35 data sets that were shared openly in 1995, PASDA's geospatial data collection has grown to include millions of data files and almost a petabyte of data. In 27 years of PASDA, users have downloaded data almost 50 million times, accessed streaming map services more than 3 billion times, and used applications such as the Imagery Navigator, Penn Pilot/Historical Imagery Viewer, Pennsylvania Atlas, and PA Mine Map Atlas, 200 million times. In 2000, PASDA was named the official geospatial data clearinghouse (now called the GIS Portal) for Pennsylvania. PASDA currently maintains data partnerships with 94 different data providers from all levels of government, non-governmental organizations, and academic institutions. PASDA also has the responsibility of creating metadata, archiving geospatial data for long term access, developing applications and tools that locate and provide access to data, creating and deploying streaming map services, responding

to user inquiries, working with agencies and data providers on data issues, and more recently, connecting to partner portals.

What is Open Data?

There are many different definitions of open data. The three major components of Open Data from the Open Data Handbook are:

- *Availability and Access*
- *Re-use and Redistribution*
- *Universal Participation*

Availability and access requires that the data be available as a whole, freely and easily accessible/downloadable. Open data are available for any use, reuse, and redistribution with no limits or restrictions. Finally, open data ensure universal participation meaning that any entity can have access to the data for any reason. For example, there is no distinction between noncommercial and commercial use (4).

Additionally, open data often refers to open source formats. This means that the data format is not proprietary and can be used in any software. The Library of Congress provides information on geospatial data formats (2).

Through PASDA and other newer geospatial data sharing portals mentioned in this report, Pennsylvania has been meeting the requirements for open data for several decades.

Why is Data Sharing Important?

The road to open data sharing has been long and at times challenging but the vibrant and growing data sharing that Pennsylvania has undertaken for decades has paid off exponentially. Free, public access to data has enabled countless initiatives to go forward. Moreover, open data frequently means significantly less expenditure of public funds when data sets are freely available. Here are just a few ways in which data sharing has benefitted Pennsylvania.

Economic Development

In the decades since Pennsylvania began sharing data openly, economic development has been one of the largest beneficiaries of access to data. Economic development projects, whether they are new manufacturing or industrial facilities, housing developments, apartment buildings, small business parks, grocery stores, restaurants, and more, all require planning and design which in turn utilizes geospatial data. An example of how free access to geospatial data provides economic benefits is the extensive use of GIS data during the Marcellus shale oil and gas boom from 2008 to 2013. Transportation, demographic, zoning, assessment, imagery, elevation, and other types of geospatial data are all used in economic development projects across our state. In addition, access to data supports logistics, supply chains, transportation, and commuting. Geospatial data allows users to analyze and visualize the planned projects and serves as a vital decision support tool.

The open data sharing environment has allowed for the growth of businesses from engineering firms, to construction and development companies, to surveying firms, technology companies, and the energy industry to name a few. It has helped Pennsylvania citizens, businesses, and governments make better, more informed decisions and grow the economy, and save time and money.

Health Care and Disease Prevention

Free public access to geospatial data in health care and disease prevention has enabled many local governments, state and federal agencies, and medical facilities to address current health care issues, visualize outcomes,

and plan services. The vital role this data plays in healthcare and disease prevention was never more apparent than during the COVID 19 pandemic. Geospatial data was accessed and used to track the disease, report its spread, identify populations in need of services such as food, oxygen, medicines, and to make vaccines more available. For more information on how GIS data and software was used to address COVID 19, please see the Pennsylvania Geospatial Coordinating Board Special Report on GIS and COVID 19 *https://www.oa.pa.gov/Programs/ Information%20Technology/Documents/ Geoboard-Report-COVID.pdf*.

Figure 1: DOQQ

Environmental Conservation and Recreation

From Pennsylvania's extensive rail trails, to the amazing parks and forests, easy and free access to geospatial data has supported recreation and environmental conservation across the state. Citizens can easily look up bike trails, boat access points, best fishing waters, state gamelands, and state park locations with a click of a button. Conservation and environmental organizations share their data easily and quickly including easements, parks, and biodiversity. Cross organizational projects and programs are easier to collaborate on and complete due to access to geospatial data. Community groups such as small watershed organizations can quickly develop an analysis of their watershed. Even books that explore our wildlife such as the 2nd Pennsylvania Breeding Bird Atlas have benefitted immensely from the plethora of available data.

Education

From K-12 to colleges and universities, every level of educational institution has embraced Geographic Information Systems (GIS) and geospatial data. Elementary students work on science projects and develop their own maps depicting their findings. Researchers and students across our public and private institutions examine climate change, environmental justice, flood risk, mining and energy resources, engineering, and business/supply chain logistics using freely accessible geospatial data.

Emergency Management and Disaster Response

The Pennsylvania Emergency Management Agency is poised to bring emergency and disaster response to a new level in the state. With the game changing NG911 initiative, emergency responders will have access to incredible new tools and data. This data will be used for everything from responding to fires, preparing for and responding to disasters like floods and tornados, and dispatching paramedics to locations where they are needed. Geospatial data is the foundation for this NG911 effort and for efforts from organizations like the Federal Emergency Management Agency, the Red Cross, and hundreds of hospitals and health care facilities around our state.

Data Sharing in Pennsylvania

As noted above, the PA DEP was the first state agency to initiate data sharing. Not only did PA DEP share data, they also provided funding for the development of PASDA. Penn State provided support for PASDA in terms of system administration, data storage, and personnel. The success of this effort led to new partners coming forward. The Pennsylvania Department of Transportation (PennDOT) was the second agency to openly share its data through PASDA beginning in 1997. Pennsylvania's Department of Conservation and Natural Resources (PA DCNR) began sharing data shortly after. These three agencies served as the initial PA Base Map creators and their data was used extensively to either supplement or serve as a framework or base layers. The Susquehanna River Basin Commission and Delaware River Basin Commission began sharing their data with PASDA in 1998. Other agencies began sharing data including the PA Department of Health, PA Fish and Boat Commission (1998), the PA Game Commission (1999), and the PA Historical and Museum Commission. More recent partners include the Pennsylvania State Police, PA Department of Agriculture, and the PA Department of Military and Veterans Affairs.

One of the first large data sharing efforts was initiated by the Pennsylvania Bureau of Geological Survey in 1998, in cooperation with PennDOT, PA DEP, and the US Geological Survey (USGS). These partners acquired the Digital Ortho Photo Quarter Quadrangles (DOQQs—black and white aerial photos of the state). In that time period, the DOQQs were provided to PASDA on hundreds of CDs and the data was then moved to FTP storage for public access. Eventually, the data was mosaicked and turned into one of the first statewide map services.

As other agencies, organizations and regional, and local governments began using GIS software and developing their own data, one major hurdle was encountered—metadata. Metadata is essentially just information about the data—what it is, when it was created, and who created it. Data was being created so quickly that metadata creation was far behind. PASDA began specialized metadata training sessions for data creators in an effort to ensure information about the data was captured, and in the hope that this outreach would encourage others to share their data. One of the first metadata training sessions by PASDA was held for Lancaster County. After completing the metadata development training funded by a grant to PASDA from the Federal Geographic Data Committee (FGDC), Lancaster County become the first county partner to openly share data in 1998 via PASDA.

In addition to providing current data, many organizations have shared historic data as well. After several years as a data partner,
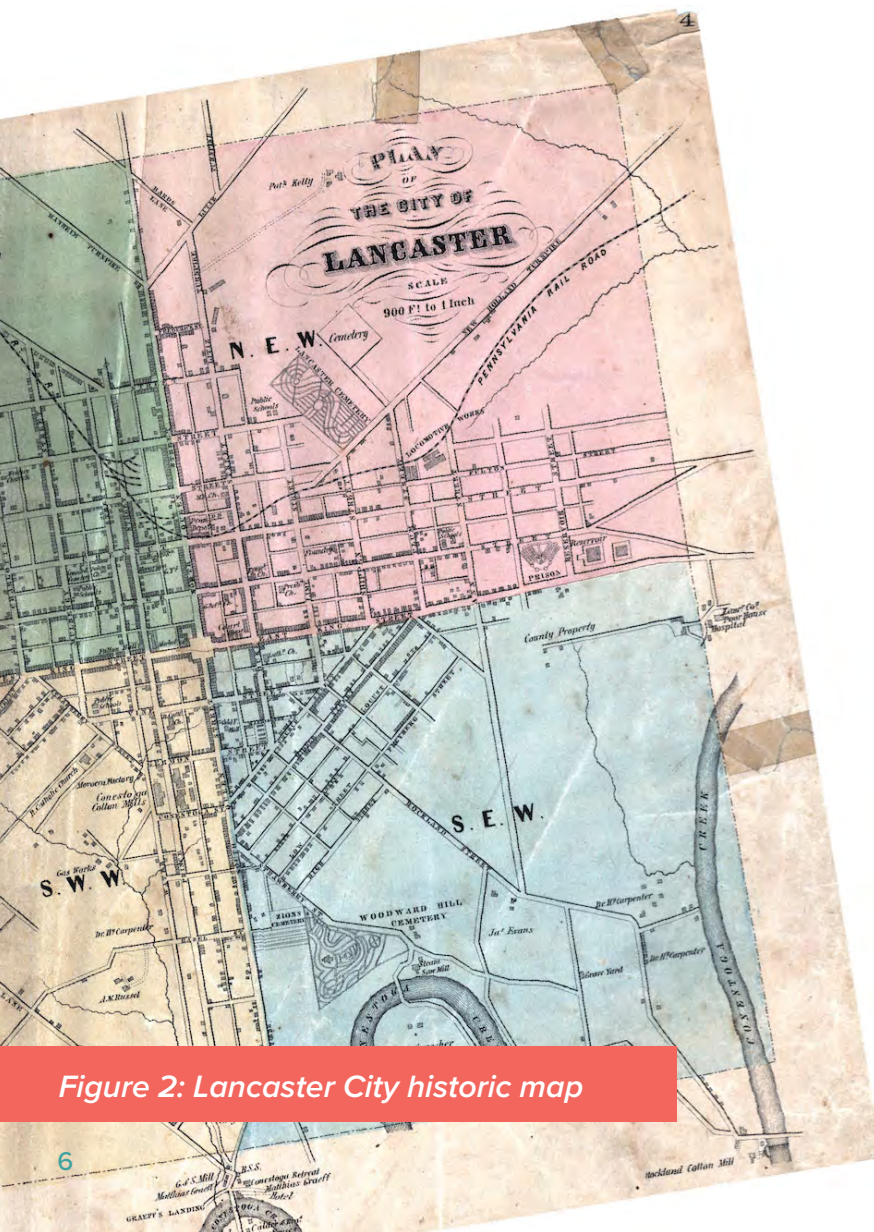

Figure 2: Lancaster City historic map

Lancaster County decided to provide access to its scanned historic county maps. These historic images are of particular importance to surveyors, developers, educators, and the county government.

Within the next several years, additional metadata training sessions were held for the Southwest Pennsylvania Commission in Pittsburgh, the City of Philadelphia, Chester County, the Delaware Valley Regional Planning Commission, and multiple state agencies, all of which resulted in new data partnerships. The City of Philadelphia was one of the first major metropolitan areas in the United States to openly share its data. PASDA worked for several years with members of the Philadelphia Mayor's Office of Information Services to develop metadata and identify data sets for sharing. The City currently shares several hundred data sets with PASDA and updates key data sets on a regular basis.

The second major metropolitan area to begin sharing data was Allegheny County. In 2008, Allegheny County began working with PASDA to provide access to its data and currently shares 57 data sets which includes data going back to 2000. In addition, Chester and York Counties became early data sharing partners with Chester beginning in 2000 and York in 2003. Mifflin County shared its aerial photography in 2001.

In addition to state and local governments, regional organizations became interested in sharing data. PASDA worked with organizations such as the Delaware Valley Regional Planning Commission (DVRPC) on a data sharing initiative that resulted in the organization sharing their

current and historic (1959, 1965-1995) imagery. Other data sharing stakeholders with a regional reach began sharing data during the mid 2000s. Two important partners are the Delaware River Basin Commission and the Susquehanna River Basin Commission who both share data via PASDA. The Lehigh Valley Planning Commission began sharing their imagery in 2010.

Non-governmental organizations (NGOs) were also utilizing GIS and developing unique data sets. One of the first NGOs to share data openly was the Natural Lands Trust which currently shares 75 data sets with PASDA. This coincided with several state initiatives including the Growing Greener initiative and the PA DCNR Rivers Conservation Program beginning in 2000. PASDA worked with PA DEP Growing Greener and PA DCNR Rivers Conservation Program to capture data being created by grantees. Current NGO partners include WeConservePA (formerly the PA Land Trust Association), Natural Lands Trust, Eastern PA Coalition for Abandoned Mine Reclamation, Heritage Conservancy, Keep Pennsylvania Beautiful, and the Western PA Conservancy.

One major outcome of sharing data is the development of unique applications. Federal, state, and local governments, business and industry, and NGOs all pull data from open data sources like PASDA to build web based applications for use internally and/ or externally by clients and stakeholders. One example of this is from WeConservePA which developed a Conserved Lands of PA application using data openly shared from PASDA. This is a common use for data served from PASDA and hundreds of organizations
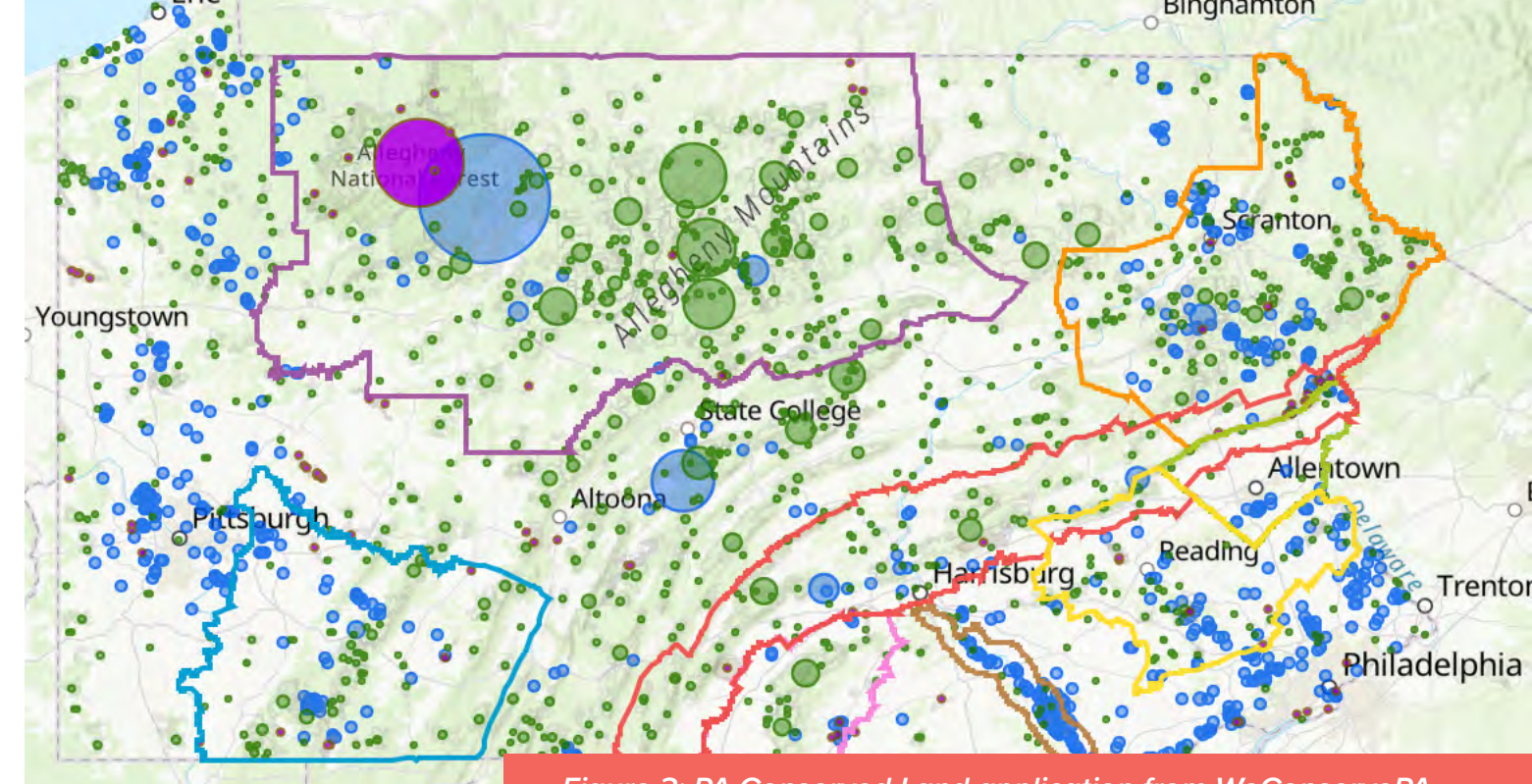

Figure 3: PA Conserved Land application from WeConservePA

take advantage of the ability to consume their own and other data in their own applications. As more data was shared and as the use of GIS increased, the need for better, updated imagery and lidar became a key component of the data landscape. An important effort from 2003 to 2006 was the PAMAP program. The goal of the PAMAP program was to create updated imagery and lidar products for the state in the expectation that local data would be shared back to the state. The data developed by this program was a vital resource for the Commonwealth and provided a meaningful base for building new data. While some local data was shared back with the state, overall it did not result in any significant local data sharing. Many local governments were still selling data and felt they could not forego the income geospatial data generated. Budget cuts and the 2008 global economic downturn ended this project. The imagery and lidar is still available from PASDA which not only provides access to the data, but also stores all the original data products in archival format. This has been a crucial activity as the data ages and software

continues to change, PASDA staff have had to refurbish and restore the data for current use. Another PA Geological Survey initiative was the PennPilot project which resulted in providing access to thousands of historic images from the late 1930s to the 1960s. PASDA took over management of the imagery in 2017. The original PennPilot application was taken offline in 2019 but PASDA created a new application to access the data. This application and data were accessed over 18 million times in 2021. In 2018, PEMA began its own imagery program in support of Next Generation 911 (NG911). Imagery was flown for the entire state in the first phase of the project. The subsequent phases will see 1/3 of the state flown each year until 2024 (5). This imagery is hosted by PASDA and was accessed almost 19 million times in 2021.

A GeoBoard Data Sharing Initiative

In 2014, the Pennsylvania Legislature passed Act 178 establishing the State Geospatial Coordinating Board (PA GeoBoard) within the Office of Administration. The first PA GeoBoard meeting was held on March 7, 2016 with a mission of providing advice and recommendations to the Governor and the public on geospatial issues, uniform data standards, coordination and efficiency in geospatial policy, and technology across different sectors (6). In 2018, the PA GeoBoard created a Data Sharing Agreement. The agreement was focused on working with local governments to acquire and share more data. At the time, data sharing by counties was still relatively limited.

Members of the PA GeoBoard leadership team, Data Task Force, Service Delivery Task Force, and PASDA came together to develop an approach to increase data sharing. With an official data sharing agreement in hand, these groups reached out to local governments, county commissioners, and others. Within a little more than two years, the data sharing landscape dramatically changed as can be seen in the coverage difference between 2018 and 2022 in the figures 5 and 6.

As of April 2022, 59 counties have signed the agreement and 38 counties are currently sharing data. One regional commission has also signed the agreement. Counties are sharing different types of data but primarily the focus has been to support all sharing and gain experience and perspective on what and how to share more broadly. Some counties are sharing dozens of data sets while others are sharing just a few. As we progress to a greater number of local governments sharing data, opportunities for aggregating and creating statewide data sets will arise. In addition, the PEMA NG911 initiative will also bolster data sharing by creating some normalized statewide data sets.

Local, Regional, or Agency Data Portals

In addition to provisioning data to PASDA, some data creators have developed their own data portals. PA DEP and PennDOT have data portals that have been active for several years. Both agencies work with and still provide data on a regular basis to PASDA so users have easy access to all data about Pennsylvania in one place and don't have to jump from portal to portal searching for information. In addition, this allows these agencies to still continue to have their data archived for long term preservation as well as access to current data via PASDA. The Commonwealth Open Data Portal provides data accessibility and exploration while showcasing web maps, story maps, dashboards, charts, and detailed data. When possible, geocoded information is added to provide a data set with enriched value for our data consumers. The Open Data Portal provides access to documents and tabular data and in July 2022 it launched its Citizens Experience Map.

Many counties have also developed their own data portals that feature unique and useful apps and tools as well as some with downloadable data. These sites are great resources for their citizens who can find parcel information, local parks and resources, historic information, and more. Many also work with PASDA and provide data to the public via PASDA's interface. Working with PASDA allows these counties to have their data curated and preserved as well as making it easier for the public to get their data in one place. A regional data portal, the Western Pennsylvania Regional Data Center, hosts tabular and spatial data for Pittsburgh and some other entities.

The Future of Data Sharing: Challenges and Opportunities

Pennsylvania's leadership in GIS data sharing over multiple decades provides multiple advantages and strategic opportunities. Along with those opportunities come some challenges that will continue to be addressed by the many organizations across the commonwealth with an interest in GIS.

Among its many accomplishments is the fact that Pennsylvania has the largest, most extensive, and most historically expansive
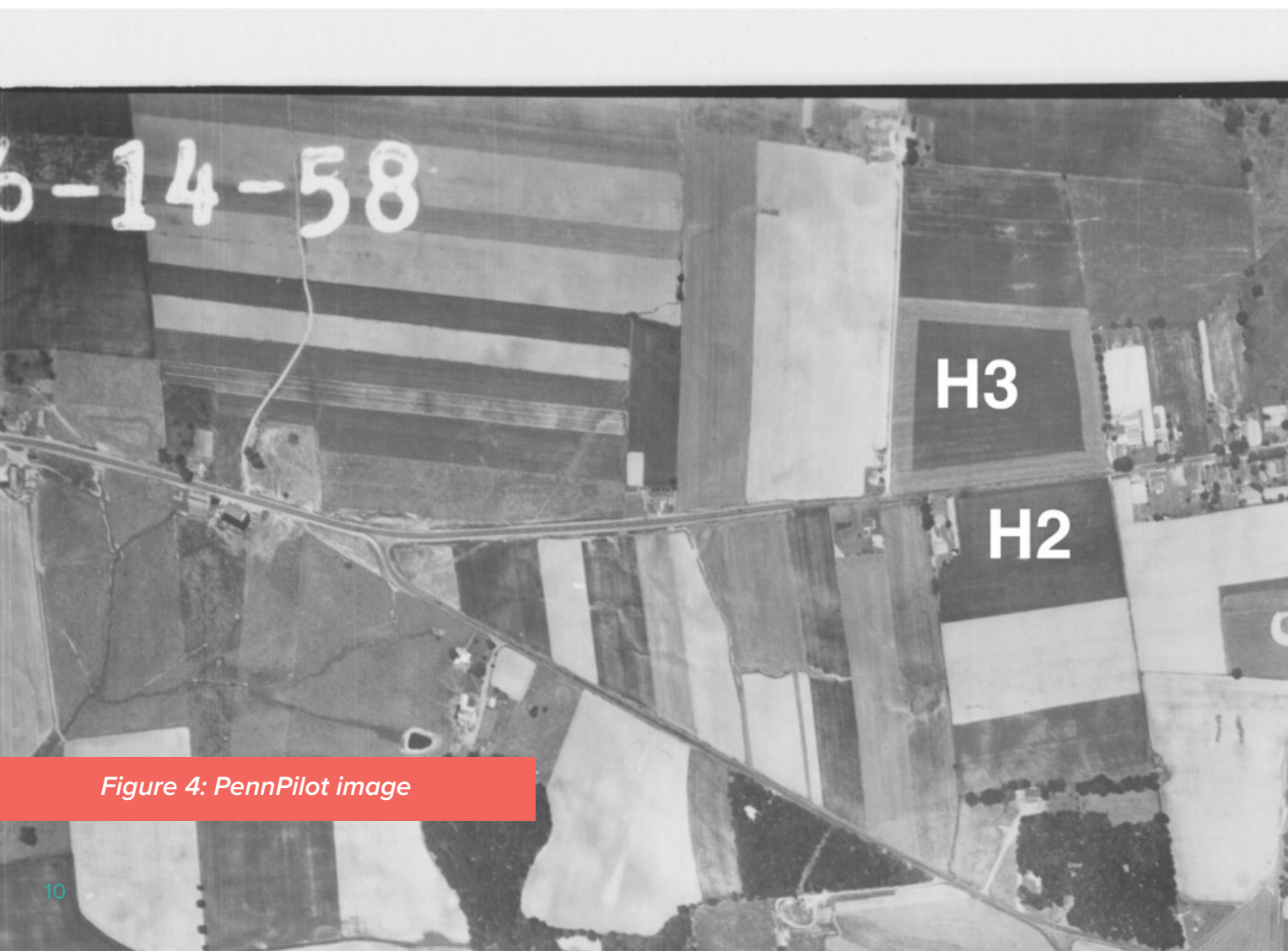
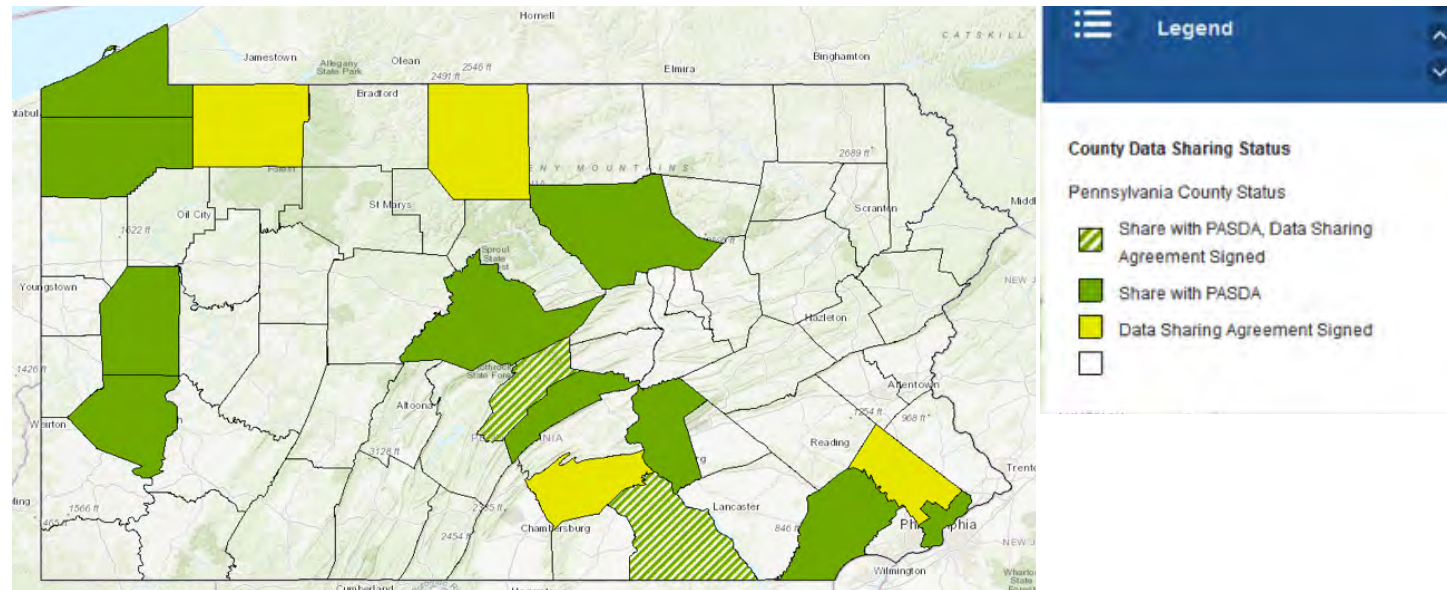Figure 5: Data sharing status in late 2019

Figure 6: Data sharing status in April 2022

publicly accessible GIS data in the country. A challenge for the future will be to continue to manage the historic data while also managing the exponential increases in the amount, size, and frequency of data. As data has become more detailed, it is also larger and more complex. Addressing so called "Big Data" (data that is large in size, complexity, frequency, attributes, and file composition) is a challenge every state faces.

The PA GeoBoard has developed a new initiative, PA Basemap 2030. The vision for Basemap 2030 is to build dense integrated 3D standardized data. PA BaseMap 2030 data will serve as the foundation to build applications, services, and will serve as a launchpad for the development of services, applications, and the building blocks for new data. This higher quality data will allow for fast, potent analytics; data-based decision making; and deeper, higher level accuracy.

The opportunity presented in this new challenge is the commitment and capability that PASDA

has demonstrated to adapt and support larger and larger data sets and synchronous data over time and serve them out to the public in new ways. This will be especially true as the PA GeoBoard PA Basemap 2030 initiative moves forward. The potential to build out services such as a 3D point cloud data service, 3D elevation data services, oblique imagery, and data visualization and processing in a browser are all currently under development and discussion. Another opportunity that comes in the form initially as a challenge is how to continue to make access to data as streamlined as possible for the user. Shared data only has value if users can access and use it easily. Having to navigate different portals to access data wastes time and money. There have been several efforts over the last 4 years to streamline data sharing and keep data up to date by directly connecting to data provider systems—whether it is a database or a data portal. PASDA has been working with several key partners including the PA Turnpike Commission and Cumberland County. The Portal to Portal initiative involves PASDA connecting via ArcGIS Enterprise's
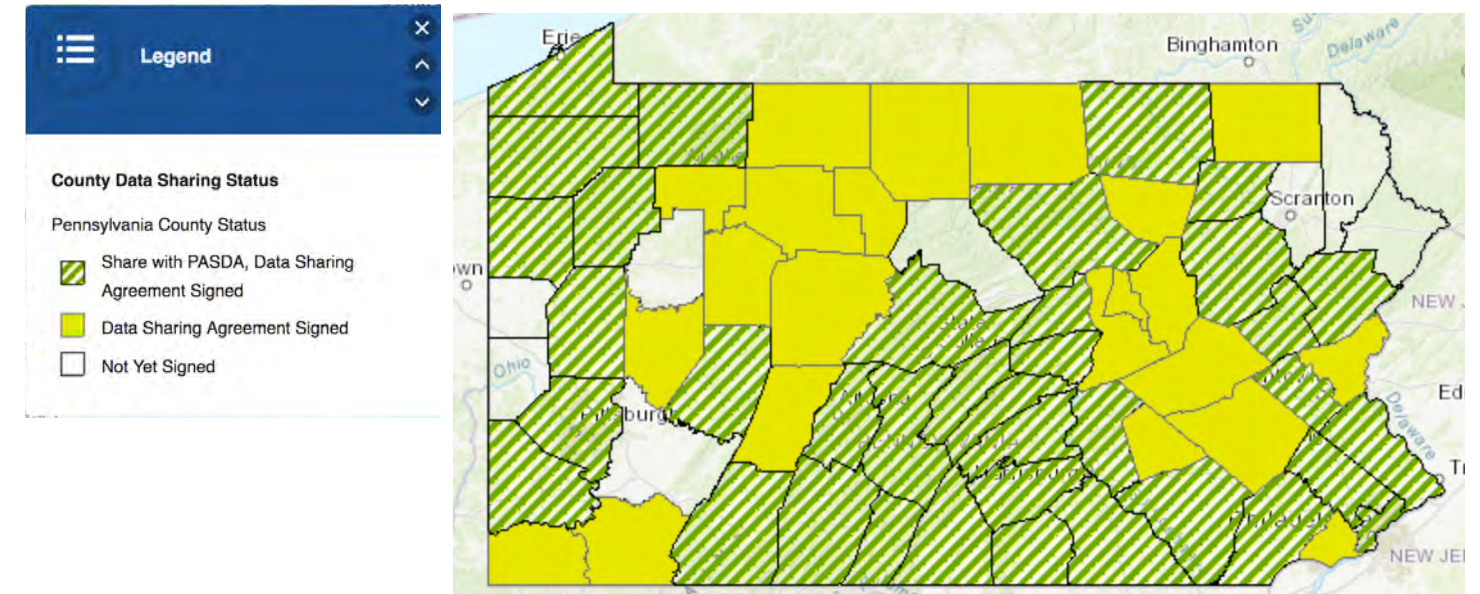
portal function to data on the provider's ArcGIS Enterprise. Both initial projects have been successful and PASDA and the Services Delivery Task Force are looking for additional partners. However, the issue is that each data provider has a unique system architecture, firewalls, and security. In Pennsylvania, there are sixty-seven different methods and timeframes for sharing and creating data at the county level. This is true across all organizations. In some cases, data is updated daily, in others monthly, bi monthly or quarterly on the data provider's end. This in turn impacts how and how often data is shared publicly. Working with partners to find additional avenues of accessing streaming data such as accessing data services and pulling them directly seems like the most viable option as we move into the future. This would allow the on the fly aggregation of PA Basemap 2030 data sets and would provide users with access to the latest data from each participating provider.

Making shared data "ready to use" is another challenge and opportunity. The GIS community is growing rapidly in Pennsylvania and includes

a diverse range of user types. From citizens looking at historic imagery of their family home; researchers exploring critical mineral availability in our state; and health care professionals analyzing how to help their patients in rural areas with lung disease get oxygen, they range from experts to GIS newcomers.

In addition to the technical challenges, there are still some remaining issues with cross jurisdictional data sharing. There have been discussions about state agencies sharing more data with local governments so that the data sharing is a more equal two-way street. Data stewardship is an ongoing challenge. A data steward is an organization that is responsible for the development and/ or updating of a particular data set. A data steward ensures that the data is authoritative. The PA GeoBoard has undertaken an initiative to develop stewards for PA Basemap 2030 data sets.

An often overlooked but vital component of data sharing is long term preservation.

There are many challenges to acquiring and providing access to data over multiple decades and preserving it for long term use. Archiving massive amounts of data is a challenge, an opportunity, and a real responsibility. PASDA has been dedicated to ensuring that data is available for future generations since its inception. The staff have long been involved in national and international initiatives and standards related to archiving and curating data including working with the Library of Congress, the Federal Geographic Data Committee Archiving, the Research Data Alliance, and providing access to data through national portals like Data.gov. Staying engaged with those organizations will require intent and investment.

Conclusion

Pennsylvania has enjoyed almost 30 years of data sharing. It is a state with a growing and diverse GIS community. The ongoing collaboration, cooperation, and good communication among those who create, use, and manage data will be vital to making this tradition continue into the future.

References

1. Kelly, M.C., Baxter, R.E., & Haupt, B.J. (2008). The evolution of spatial data infrastructure (SDI) geospatial data clearinghouses: Architecture and services. In Encyclopedia of Data Warehousing and Mining, (2nd ed.). Hershey, Pennsylvania: IGI Global.

2. Library of Congress, Recommended Formats Statement, https://www.loc.gov/preservation/resources/rfs/geo-carto.html.

3. National Research Council, Mapping Science Committee. (1993). Toward a coordinated spatial data infra-structure for the nation. Washington, D.C.: National Academy Press.

4. Open Data Handbook, http://opendatahandbook.org/guide/en/what-is-open-data/

5. Pennsylvania Geospatial Coordinating Board PEMA Statewide Imagery Project Report, August 2021 https://www.loc.gov/preservation/resources/rfs/geo-carto.html

6. Pennsylvania State Geospatial Coordinating Board, Annual Report 2016
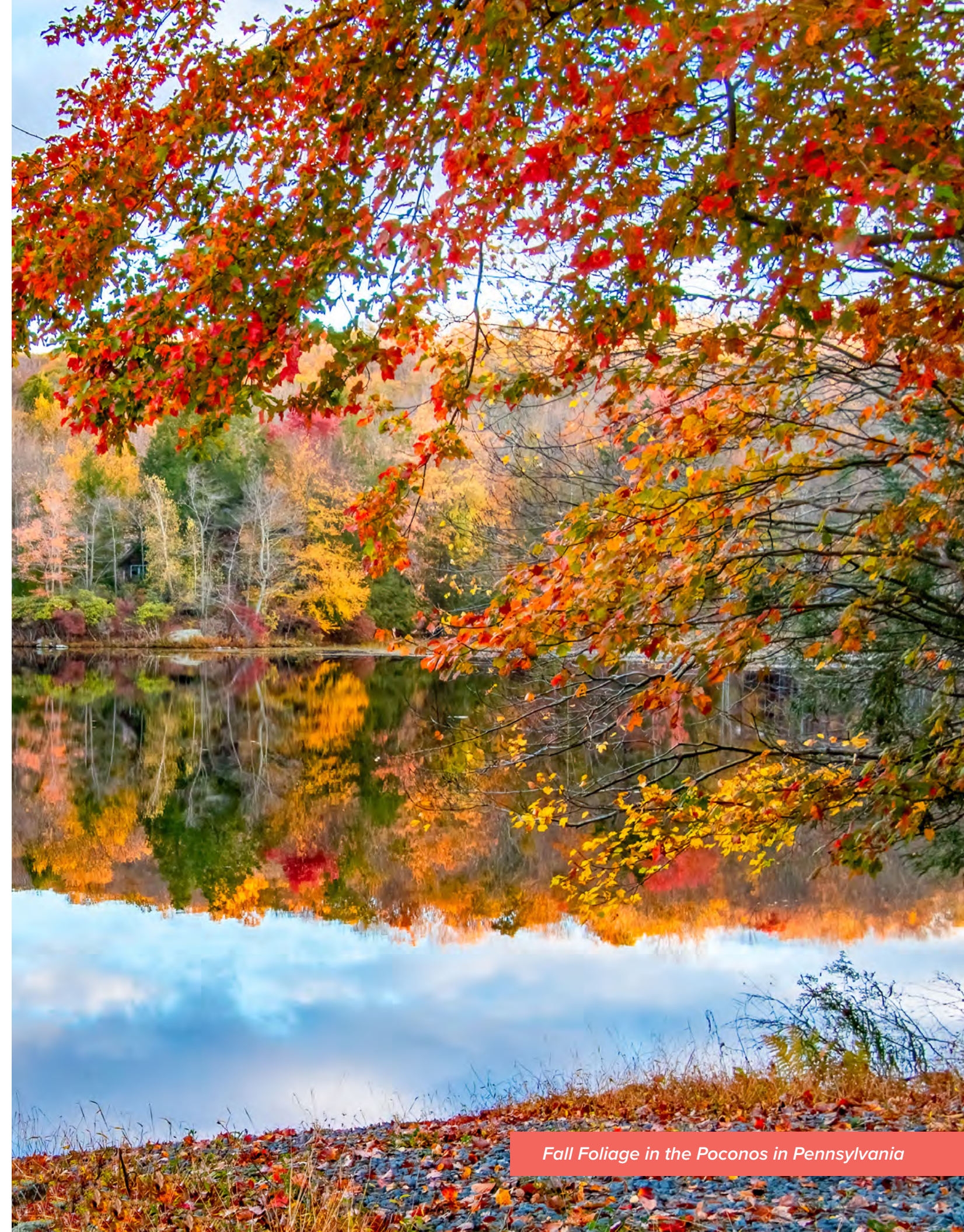
Fall Foliage in the Poconos in Pennsylvania