

Exploring Geospatial Options for Information Discovery and Access:

The case of Wisconsin's historical newspapers

Mark A. Rozmarynowski

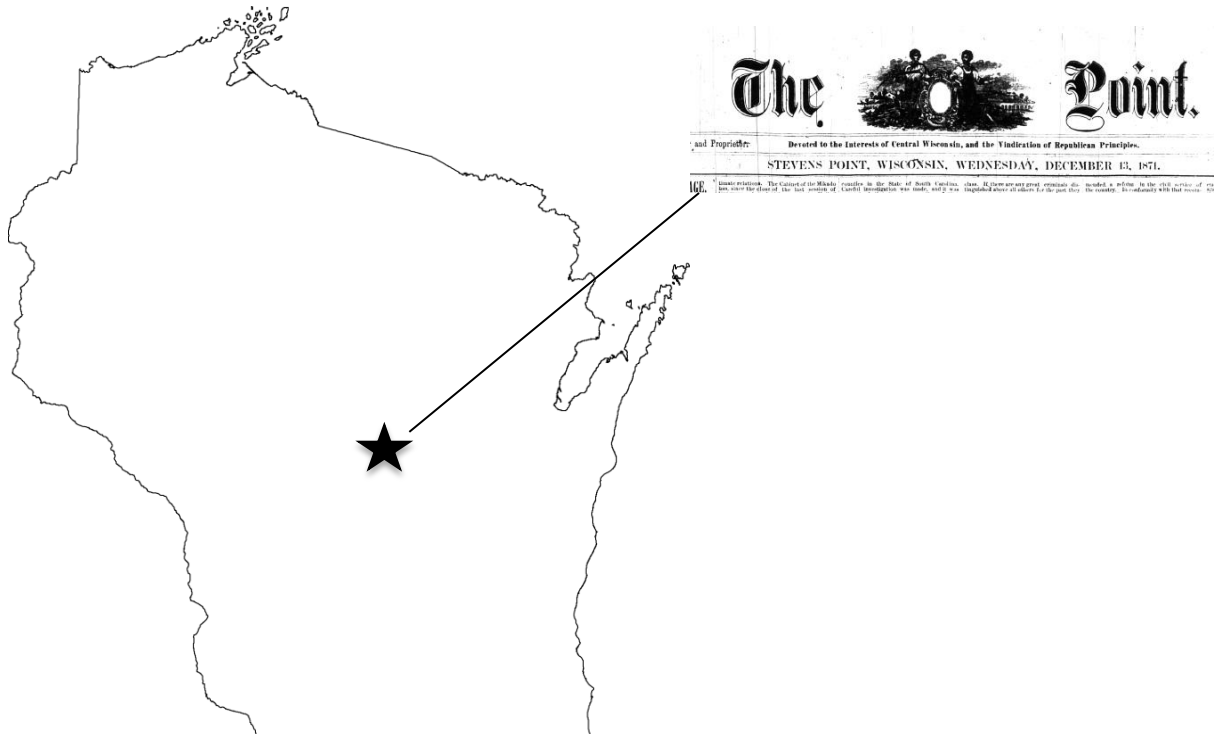
Masters in Geographic Information Systems

Pennsylvania State University

Capstone Project Paper

Advisor: Deryck Holdsworth

12/01/2014



Contents

Abstract	4
Introduction	5
Geographic Information Systems	10
Project Background.....	13
1. Ancient Land and First Peoples, 13,000 y.a.-present.....	13
2. Fur trade era, 1650s-1850s	15
3. Territorial era, 1787-1848.....	16
4. 19 th century immigration	16
The Geographies of Newspapers	18
The Challenge of Using Historical Newspapers	21
Digital Newspaper Collections	22
The National Digital Newspaper Program (NDNP)	24
OCLC's WorldCat	25
Stanford's Rural West Initiative: Journalism's Voyage West	26
Capstone project prototypes	28
Project Challenges.....	36
Definition	36
Uniqueness.....	36
Conflicting Sources.....	38
Selecting Attribute Data.....	40
Verification.....	41
Incorporating Temporal Data.....	42
Determining the "Geography" of a newspaper?	43
Local Newspaper Holdings.....	44
Prototype Evaluation	48
Conclusions	49
References	51
Appendix A.....	54
Appendix B	55

Abstract

Historical newspapers are a major source of information for a given place and time. The vast majority of newspaper content, amounting to millions of pages, is still scattered in print collections such as libraries, archives, and historical societies around the world. Over 3,500 newspaper titles have been published in Wisconsin to date. Most contemporary text-based library research tools provide limited options for helping users discover and access this content. Using a nine-county area in central Wisconsin as a pilot study area, this project examines the potential for a web-based GIS to help facilitate discovery and access of geo-referenced historical United States newspaper content currently only available in traditional physical formats.

Keywords: GIS, Geographical Information Systems, History, Wisconsin, Newspapers

Introduction

There are many tools available to academic librarians to help their students, faculty, staff, and community members with research. Some of these tools, such as internet search engines, are still relatively new. Others, such as library catalogs, have been in use for hundreds or even thousands of years. This project paper investigates the possible development of a new tool for the arsenal, one for helping library users identify and locate historical newspapers. While this capstone project focuses on developing a discovery tool for titles found in a small area of central Wisconsin, it is hoped that this information will be applicable in other regions, or even on a national level. And while this study focuses on print copies of these titles, this work may also serve to lay a foundation for facilitating access to digitized content as well.

To begin, academic librarians serve patrons at their campus libraries in two key ways that may be broadly summarized as *discovery* and *access*. By *Discovery* is meant assisting library users with the process of identifying potential sources of information (books, articles, etc.) that may satisfy an information need. By *Access* is meant assisting the user with locating and obtaining a usable copy of that information source, either in print or online, from local collections or locations around the world.

A variety of research tools aid contemporary academic librarians with the processes of discovery and access, including online resources such as internet search engines, online library catalogs, and digital repositories, as well as journal, magazine, and newspaper databases (Katz, 2002). The rapid development of the Internet and the World Wide Web has enabled contemporary sources of information to be both discoverable and accessible entirely in an online environment. Current newspaper content, for example, is usually available in a digital format and indexes and databases for locating and accessing this content have been available online since the early 1990s.

The abundance of easily accessible information online may deceive users into believing that “It’s all on the web.” But writing has been in use for about 5,000 years, and digitization efforts are still (relatively speaking) in their infancy, so a tremendous quantity of potentially useful historical information is still locked up in traditional print formats. Much of this traditional content is housed in libraries and archival collections, and is often only accessible on-site. Historical newspaper content is a prime example of this content.

An estimated 250,000 newspaper titles may have been published in the United States since the first paper was started at Boston in 1690 (Library of Congress, 2013). The newspaper industry arrived in the United States from Europe about 300 years prior to the development of the internet and the World Wide Web, and the number of publications in print in the United States peaked around 1915 (Clark & McGee, 2012). Each newspaper title produced dozens (in the case of rural weeklies) or even hundreds of issues (in the case of urban dailies) annually. Individual issues have ranged in length from four pages in the very earliest papers to dozens of pages in multiple sections for later titles, so during the course of its lifetime a single newspaper may produce thousands of pages of content. These thousands of pages will include tens-of-thousands of articles, editorials, letters, personal and legal announcements, and advertisements (Oehlerts, 1958).

This abundant content serves as a *de facto* record of human culture for communities around the country. The “local” nature of a newspaper makes newspapers an ideal source for researching a topic in history at both a given *time* and *location*. Unfortunately, facilitating discovery and access for historical newspaper content usually means patiently navigating a mix of traditional print directories, online indexes and library catalogs, scattered library and historical society websites, and dusty local history tomes. Moreover, the various resources developed to assist users in locating newspapers (or other) content tend to provide limited options for exploring content geographically. Contemporary online

databases, for example, tend to provide search options that are based on traditional print access points, like title, author, or subject headings. Keyword searching is typically available in full-text digital resources, with additional limiters (e.g., publication date, physical format, or publication language) available to help narrow results. But in these resources most of the options for discovering information using *location* as a search parameter are limited.

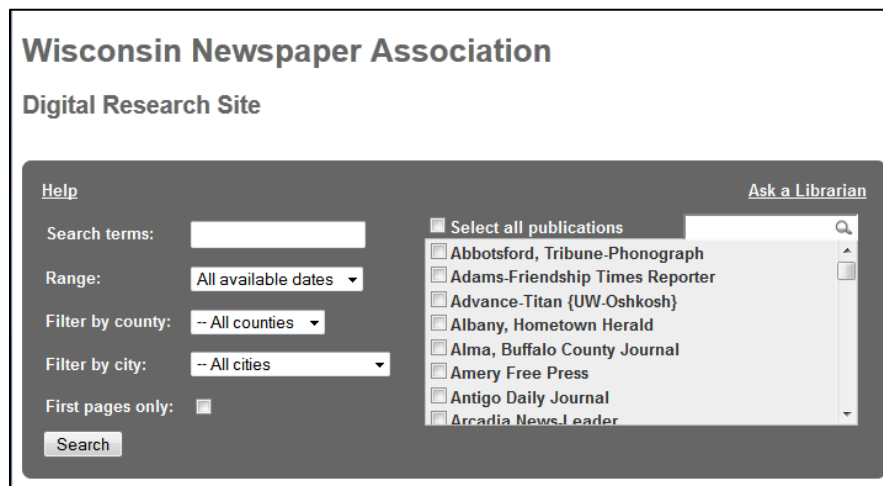


Figure 1. A search interface for locating digital newspaper content. Retrieved from the Wisconsin Newspaper Association's Digital Research Site. Note the limiter to "filter by county."

The lack of viable geographic discovery tools may result from the fact that search options in library resources tend to be text-based, relying on place-names, or toponyms, as points of access, rather than geographic coordinates. In order to search for information about a specific location a user types in the name of the location as a string of text, and relevant hits are identified algorithmically, matching these strings of text with those indexed in the database. Text-based location searches can be problematic in various ways, for example when place names are not unique (e.g., *The Simpson's* hometown of "Springfield") or when there have been variations in the spelling of a place name (e.g.,

Milwaukie vs. Milwaukee) over time (Beall, 2010).

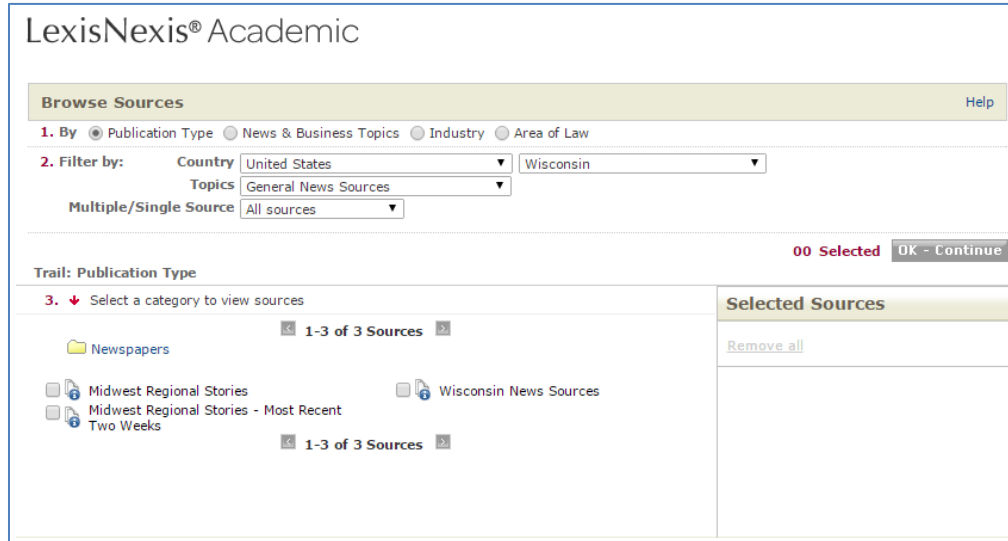


Figure 2. In online databases, browsing by location usually involves clicking through text-based menus.

Note the “Country” drop-down menu in this example retrieved from LexisNexis Academic.

Browsing by location is another typical text-based search option. In order to browse by location a user clicks through successively more refined levels of political geographies, starting with the name of a country, then a state within that country, then a county or borough, and finally down to a city or town. This can be problematic for many reasons, such as in the case of geographies with undefined or historically shifting boundaries. Additionally, rural areas and natural features are often not represented in such a hierarchy. These options rely on political boundaries, which may change names, move, split, merge, etc. The text-based search interfaces described above also presume a level of familiarity with an area of interest, and to use them a researcher must know in advance the state, county, and town of interest, including both current and past place names, as well as the current and past boundaries of their associated geographies.

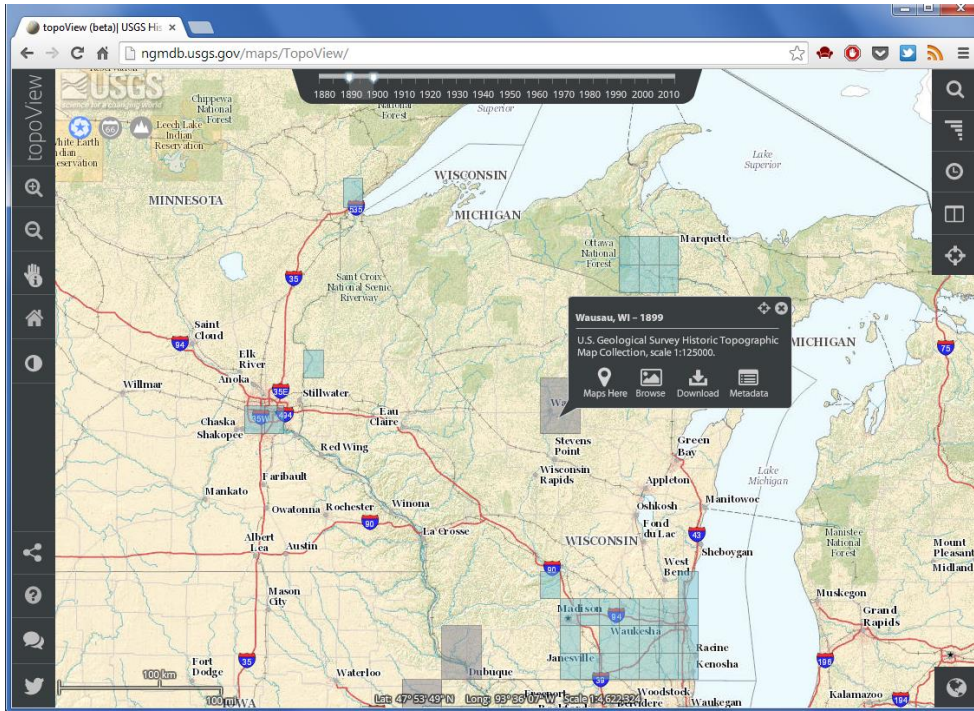


Figure 3. This map-based interface provides a visual, geographic search option for identifying historical topographic maps. Retrieved from USGS topoView (beta).

By way of comparison, a geographic information retrieval tool, such as the one in figure 3 provides a visual, map-based, interactive approach for discovering and accessing information using location as a parameter. To locate materials a user clicks on a point on a map, or selects a geographic area (polygon) of interest. This allows a user to navigate a geographic region visually and to identify areas of interest without knowing place names or boundaries in advance.

This kind of interface has been used in many projects that providing access to information sources with obvious geographic boundaries, such as historical maps or aerial imagery. Work done in developing these projects has established a number of standards and practices for geographic search that might now be applied to systems that would enable searching for information with less clearly defined geographic boundaries. A newspaper is an example of a document that at first blush would

seems to have an obvious geography, such as the town or county where it was published, but whose geography can turn out to be a bit fuzzy, as will be further discussed below.

The lack of truly sophisticated geographic search options in library resources persists, in spite of at least two decades of related research, and despite the fact that location matters for people seeking information. A 2006 study of Yahoo.com search engine queries found that 12.7 percent of searches included place names (Jones, Zhang, Rey, Jhala, & Stipp, 2006). Other studies have suggested that percentage may now be even higher, especially as search engines have adopted more sophisticated geographic retrieval algorithms and users have raised their expectations as a result. In sum, essentially, along with *who*, *what*, and *when*, people want to know *where* things are happening.

Geographic Information Systems

Finding out *where* something is happening (or has happened) is no doubt easier today than a hundred years ago thanks to the development of the Internet and the profusion of new web-mapping services. Web maps are essentially online geographic discovery tools, and they are made possible thanks to the continued development of underlying Geographic Information Systems (GIS). A Geographic Information System may be described as a combination of computer hardware, software, and data that supports the collection, organization, dissemination, production, and analysis of information with a spatial or geographic aspect. In short, a GIS can manipulate any data whose *location* can be defined.

The rapid development of Geographic Information Systems since the 1960s has resulted in a variety of new ways for exploring *geo-referenced* data. Even if they are not recognized as such, Geographic Information Systems are used in a number of day-to-day contexts, such as websites for mapping travel routes, visualizing election results, predicting weather patterns, managing urban development, or modeling the spread of diseases.

Historical geographers have adopted the tools and techniques of Geographic Information Systems to help further our understanding of where things happened in the past and to re-examine historical events in their geographic context (Knowles, 2008). Historical GIS (HGIS) has been an area of research since at least the late 1990s (Gregory & Healey, 2007). Recent historical GIS work has examined such diverse topics as environmental, religious, and social history (Bonnell & Fortin, 2013), as well as agricultural, political, and racial history (Gregory & Geddes, 2014), among others.

Historical GIS research requires the availability of GIS datasets, such as the historical census and population data hosted by the National Historical Geographic Information System (Minnesota Population Center, 2011). The NHGIS includes both the data from past censuses, such as population figures, and their associated geography, such as historical county boundaries. The datasets help provide both geographic and temporal context for historical studies, but a theme commonly repeated in HGIS research is the lack of historical GIS data, and the time-consuming nature of building a GIS database from scratch (Gregory & Healey, 2007).

Geographic Information Retrieval (GIR) is a branch of Information Science that has built on the foundations of GIS to develop a variety of tools and methods for enhancing information online discovery systems, using location as a primary point of access (Purves & Jones, 2011). As noted above, Geographic Information Retrieval studies have found that information seeking behavior routinely involves at least three main aspects: thematic (subject matter), temporal (time), and spatial (location) (Palacio, Cabanac, Sallaberry, & Hubert, 2011). These three aspects may be summarized as *what*, *when*, and *where*. Larson also noted that *place* is a key facet for many researchers, along with *people*, *subjects*, and *time periods* (1996). GIR research has typically been done in the context of digital libraries composed of digitized text or born-digital content. For example, GIR research has explored automated methods of identifying and disambiguating toponyms for use in geo-referencing, evaluated methods for parsing out, identifying,

locating, and analyzing geographic references in unstructured digital texts (Purves & Jones, 2011). Significantly less work has been done on geo-referencing physical library or archival materials like the work done to georeference archival data in the United Kingdom's National Archives (Clough et al, 2011).

But, many (if not most) physical collections of materials in libraries and archives have been cataloged, and are searchable online, and include at least *some* geographically-enriched descriptive metadata which might enable geographic discovery. Buckland et al. explored the possibility of linking library catalogs with digital gazetteers in order to facilitate geographic search via a web-based interface with mixed results (2007).

It is quite possible that the work done on Geographic Information Retrieval in a digital library context could be applied to physical collections, and vice-versa. Much of the content in digital libraries has been digitized from physical collections with descriptive metadata added to enhance discovery. The metadata is often derived from, and therefore may be linked back to, original library catalog records. This is a complex issue, that will take much time and effort to sort out and this project is only a small piece of that puzzle.

The concept of a geospatially enabled historical newspaper database and discovery tool intersects various strands of Geographic Information Systems, Historical Geography, and Geographic Information Retrieval. It is not difficult to imagine ways that the availability of a geospatially enabled newspaper database and online discovery tool would enable the exploration of a variety of historical topics in disciplines such as economics, urban development, medicine, or natural resources management. One might imagine identifying and searching all of the newspaper titles that were in publication when the Wisconsin River was first dammed for papermaking; or looking for information about local indigenous cultures in the first newspapers published within fifty-miles of a newly

discovered archeological site. Geolocating information tends to provide many new and possibly unforeseen options for exploring information (Gregory & Healey, 2007).

The discussion so far has been presented in the context of print newspaper collections, but developing a geospatially enabled database and online discovery tool for print collections now may also help lay the groundwork for navigating digitized content in the future. There are a number of digitization projects beyond the Library of Congress collections that have been completed or are now underway, amounting to millions of pages of digitized content. In the process of developing a geographic discovery tool, a number of complex issues related to the collection and organization of newspaper data and geolocation would be ironed out. But first, to provide context for this project, some historical background on Wisconsin and newspapers may be in order.

Project Background

For this capstone paper a pilot study was conducted using nineteenth and twentieth century newspapers from communities in a several-county region in central Wisconsin. For those unfamiliar, the State of Wisconsin is situated geographically in the northern United States, at the western end of the Great Lakes. The state is rich in natural resources, including many lakes and forests in the north, and prairie and agricultural land in the south, with many rivers running throughout. The Wisconsin Historical Society divides the state's history into several somewhat overlapping eras (WHS, 2014). Each of these historical eras is represented to a degree by corresponding developments in the newspaper industry:

1. Ancient Land and First Peoples, 13,000 y.a.-present

The Wisconsin area was home to several Native American tribes when the first European explorer landed in 1634 near present-day Green Bay. The first American newspaper would not be published until some fifty years later at Boston (*Publick Occurrences...*), and it would be another one-

hundred and fifty years after that before the first Native American newspaper (*Cherokee Phoenix*) was founded in Georgia (Clark & McGee, 2012). That first Native American newspaper was founded five years before the first paper of any kind was published in Wisconsin. Yet, even this earliest period of Wisconsin history (or pre-history) is represented by newspapers to an extent. The descendants of the First Peoples began using the press to tell their story in the 1960s. Today almost every tribe runs a newspaper, including the Ho Chunk, Menominee, Oneida, and Ojibwa (or Chippewa). Those newspapers are headquartered on the tribal reservations found throughout the state.



Figure 4. The locations of contemporary Indian reservations and related tribal newspapers in Wisconsin. Basemap by Great Lakes Intertribal Council, Inc.

2. Fur trade era, 1650s-1850s

The French and British fur trades dominated the first two-hundred years of Wisconsin's recorded history. Settlements during this period were largely clustered around trading posts and military forts (Nesbit, 1973). Green Bay (then known as Navarino) was one of the earliest settlements and it was there in 1833 that Wisconsin's first newspaper, the *Green Bay Intelligencer*, was founded (Oehlerts, 1958). The arrival of newspapers comes near the end of the fur trade era, as settlement and agriculture are promoted by the new United States government as a means to replace the economy of fur.

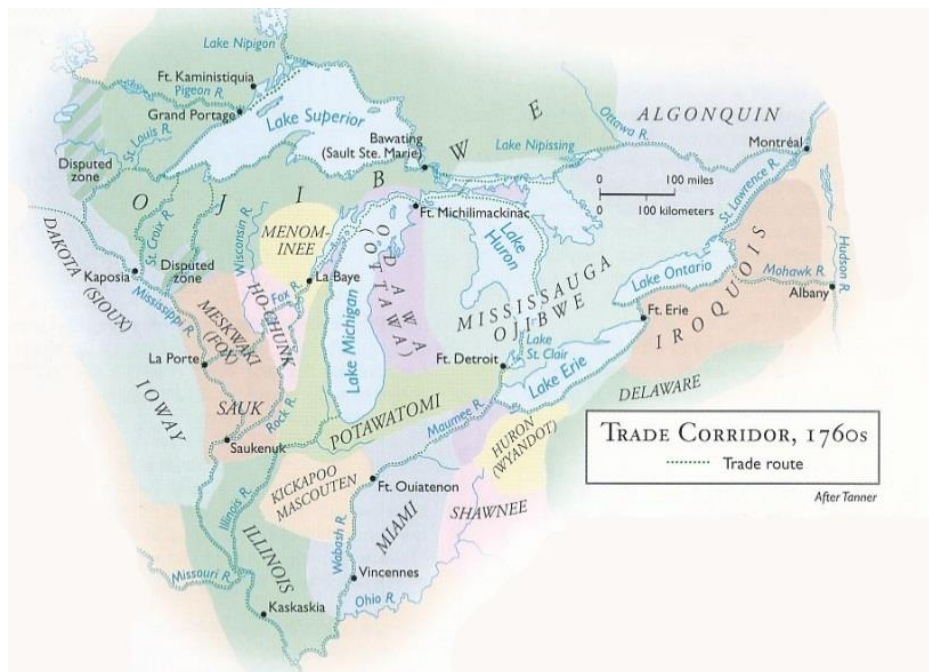


Figure 5. Trading routes of the fur era. The first newspaper in Wisconsin would be founded on Green Bay (La Baye) along on the Fox-Wisconsin River route. Map by the Wisconsin Cartographer's Guild.

Within a decade of the founding of *Intelligencer* several other papers sprang up, first in Green Bay, then two in Milwaukee, also the site of a trading post. Mineral Point was a small town located at the heart of the state's early lead mining industry since the 1820s, and that town produced the *Miner's Free Press* in 1837. The towns of Racine and Madison, the latter of which was eventually to become the site of the state's capitol and second-largest city, had both founded newspapers by 1838.

3. Territorial era, 1787-1848

As the fur trade declined, the Federal government began auctioning off land in the Wisconsin Territory and promoting its development. Many settlers arrived during this period from the eastern states, particularly New York, as well as from Europe (Nesbit, 1973). A few aspiring Yankees appear to have brought presses with them. Swiss, Norwegian, and German immigrants also published some of Wisconsin's earliest newspapers, and these in their native languages. The *Wisconsin Banner*, for example, was the state's first German language paper founded in Milwaukee in 1844. Soon after, *De Nieuwsbode* and *Nordlyset* were the first Dutch and Norwegian newspapers started in Wisconsin.



Figure 6. Ethnic or language newspapers appear to have served a dual-role of helping new citizens participate in American society while maintaining their own cultures (Oehlerts, 1958).

4. 19th century immigration

By 1848 Wisconsin had attracted a large enough population to become the newest state in the union and economic development continued. Many newspapers during this period appear to have been founded primarily to promote the interests of an area's largest industry, such as logging or mining.

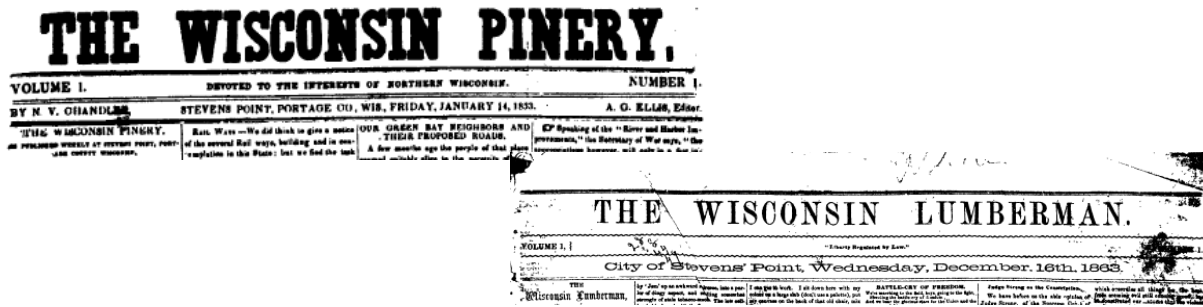


Figure 7. In Stevens Point, WI the first two local newspapers, *The Wisconsin Pinery* and *Wisconsin Lumberman*, reflect the influence of the logging industry during the 19th century.

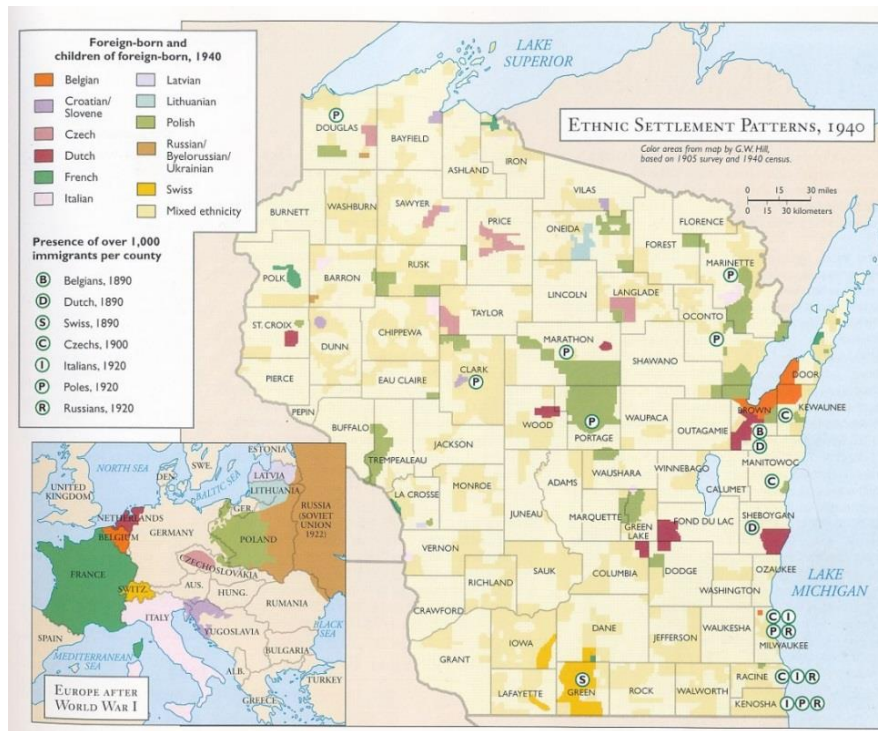


Figure 8. Immigrant settlement patterns, 1940. Note the large Polish populations in Central Wisconsin, denoted with a (P), which produced several Polish language papers. Map by the Wisconsin State Cartographer's Guild.

During the later 19th century the waves of immigration continued bringing a large Polish population to central Wisconsin in the Stevens Point, Portage County area. A number of Polish language newspapers were founded during this era, including *Rolnick*, the Polish word for Plowman. That

newspaper's name, *Rolnick* demonstrates both the influence of the Polish community and the growing importance of the agricultural industry during this period.



Figure 9. “Rolnik”: the first of several Polish language newspapers published in Stevens Point, WI, USA.

These new immigrants brought with them new attitudes about American slavery and capitalism. The *American Freeman* in southeastern Wisconsin represented the abolitionist movement during this era, while the local *Portage County Gazette* openly sympathized with the labor movement in their first editorial statement.

One dramatic, and instructive, example of local newspapers mirroring the events of the greater society involves German language newspapers in Wisconsin. During the nineteenth century German immigration eclipsed all other ethnic groups arriving in Wisconsin and many German language papers started up to serve this new population. They flourished for decades, and yet most of the dozens of German language papers in the state met a sudden rapid demise as a result of anti-German sentiment and suspicion during the First World War (Oehlerts, 1958).

The Geographies of Newspapers

The history of early newspapers in Wisconsin is summarized in a directory of newspapers compiled by Donald Oehlerts of the Wisconsin Historical Society. The content within the directory is organized by place of publication, first at the county level, then within each county by city or town (Oehlerts, 1958). Two things about this newspaper directory formed the basis for this project:

First, there have been far more newspapers published in the state than one might expect, at least 3,500 titles to date in a state with just seventy-two counties and under a dozen cities with a population of over 50,000 persons (WDSC, 2013). This indicates that the state's newspaper industry has been quite volatile, with many papers coming and going in relatively short periods of time. For researchers, discovering the existence of these papers is difficult if not impossible given the limitations of current search tools. Getting a hold of copies adds a whole other layer of difficulty.

Second, the author chose to organize his newspaper directory by place of publication, which implies that newspapers may have geography, or possibly several *geographies*. It seems obvious that most "local" papers have a limited geographic area of news coverage, and most also typically have a limited geographic area of distribution. These limited geographies may have been a necessity in the past, as travel options and communication technologies were somewhat limited. But even in the modern era of syndicated content and virtually instant communications the tradition of local papers providing local news coverage persists.

The commonly used phrase "local newspaper" implies that the information published in a given title will be mostly about, or at least somehow relevant to, people and events at a particular locale. Quite often, the title of a newspaper may give a sense of that geographic area of interest. In Stevens Point, Portage County, Wisconsin, for example, there is a daily newspaper called the *Stevens Point Journal*, named for the City of Stevens Point. News coverage in the *Stevens Point Journal* tends to focus on events within the city itself as well as the surrounding areas, including the smaller local towns and villages such as Plover and Amherst, or rural locations within Portage County. State of Wisconsin, national/U.S., and international news are also provided, but the paper's primary scope and focus of reportage is the Stevens Point area.

There is also a weekly paper published in the city, named *The Portage County Gazette*, and another weekly, *Central Wisconsin Sunday*. Statewide there is also a daily *Wisconsin State Journal*, and Wisconsin's major metro newspaper, *Milwaukee Journal Sentinel*. Each paper writes with a perspective that is tailored to its apparent level of geography, the exception possibly being the Milwaukee paper which provides both city and statewide news coverage.

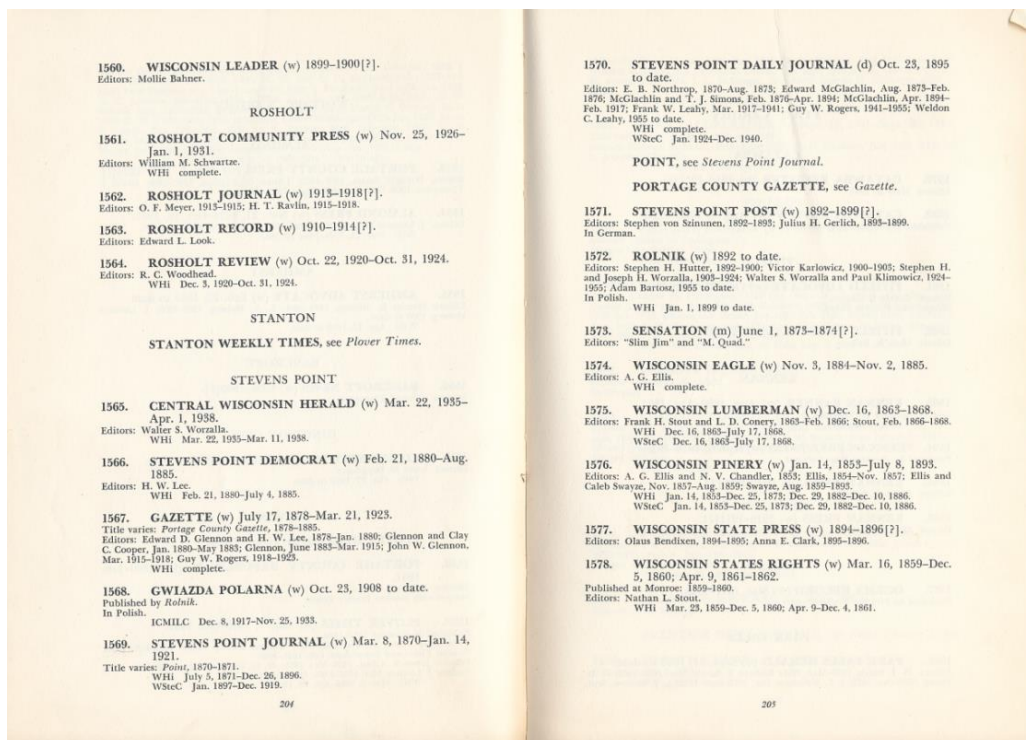


Figure 10. Two pages from Oehlerts' 1958 newspaper directory for Portage County, WI, including the towns of Rosholt, Stanton (*Plover*), and Stevens Point.

The geographies that newspapers represent can make historical research quite convenient at times. If you want to know what was going on in a particular place at a particular time in the past, you can consult a newspaper published in or near that place during that time. Not only are public events covered in local papers, but private ones as well such as weddings, funerals, graduations, and so forth.

The historical record is quite rich, assuming one can identify and obtain newspaper content for a given time and place. Unfortunately, this can turn out to be quite challenging.

The Challenge of Using Historical Newspapers

Part of the challenge of using newspapers in research is related to the dynamic environment that the industry has occupied over the years. As noted, at least 3,500 titles have been published in Wisconsin, and a U.S. Library of Congress research project estimated that there may have been as many as 250,000 newspaper titles published nationwide (--, 1987). Many of those titles folded in a short period of time, and many more have changed names over the years, some several times. Other papers have merged, or have been bought or sold, often numerous times, so generally speaking no one title contains a complete record of all titles.

An additional challenge is that no single research tool appears to provide a complete listing of all of these titles, and that includes the massive online directory collected and maintained by the Library of Congress. So students, historians, researchers, and genealogists have to piece together information about these publications from a variety of online and print indexes, newspaper directories, bibliographies and local histories, or in consultation with local experts.

It goes without saying then that no single collection includes the content from all of these titles either. In fact, some papers no longer appear exist except for a brief mention in the historical record. Libraries have tried to preserve as many of these documents as possible, but as library budgets fluctuated over the decades, subscriptions were dropped, storage needs outgrew physical spaces, and many libraries have had to discard older collections, or look for alternative methods of preservation, such as microfilm.

Digital Newspaper Collections

Since the birth of the Internet and World Wide Web, various initiatives, both public and private, have attempted to digitize and consolidate newspapers collections online, either freely or for a fee. Anyone that has dipped into their family history is likely familiar with popular subscription databases such as Ancestry.com or NewspaperArchive.com. But the sheer quantity of content has played a limiting factor in these efforts. The financial costs involved in identifying, locating, obtaining, scanning, and hosting all of the estimated millions of pages of newspaper content would be enormous.

Even as the cost of technology used for digitizing content has declined, and the speed of digitization equipment and workflows has improved, the spread of available content at disparate locations around the globe would make it difficult if not impossible to obtain titles for purposes of scanning.

As a result many digitization initiatives have focused on a limited time period, a limited locale, or specific type of publication. Coverage in these collections is often incomplete and runs of titles can be spotty at best. Since much of the digitization work is done not from original paper copies but from rolls of preservation microfilm, other problems can creep in, including lost or blurred pages, or a generally poor quality, black-and-white digital copy of the finished product.

For online sites that do have digitized content, the options for exploring collections by location are somewhat limited. Typical database search options are limited to keywords or text-based location searching, which may work well if you live in a town with a unique name, such as “Nekoosa,” but can be a bit problematic if like the fictional cartoon character Homer Simpson you live in a town with a more commonly occurring name like “Springfield.”

The option to “browse by location” has also recently become available in some online collections, but this simply means clicking through successive hyperlinks in order to navigate through progressively more refined regions, from country to state to county, and so on. For this to be a truly useful means of exploring by geography, a researcher would likely already have to know quite a bit about the location of interest, including current and historical place names, or the county a particular town resides in. This “breadcrumb trail” navigation also provides limited options for places with multiple historical names, for those that may overlap borders, or for exploring rural or natural areas that are not identifiable with obvious political hierarchies.



Figure 11. Screen capture from NewspaperArchive.com, showing a browse by map option. The map only functions at the state level.

You Are Here: [Home](#) > [US Newspaper Lists](#)

LIST OF US NEWSPAPERS BY STATE

Explore Historical Newspaper Articles from US Newspapers

Select State (52)	Select City (192)	Newspaper (6)	Year	Month
Rhode Island (1)	Ripon (5)	Stevens Point Daily Journal (91)		
South Carolina (4)	River Falls (1)	Stevens Point Gazette (19)		
South Dakota (3)	Sauk City (2)	Stevens Point Gazette And Stevens Point Journal (1)		
Tennessee (2)	Shawano (3)	Stevens Point Wisconsin Lumberman (4)		
Texas (94)	Sheboygan (14)	Stevens Point Wisconsin Pinery (12)		
Utah (8)	Sheboygan Falls (3)	Wisconsin Pinery (4)		
Vermont (1)	Shullsburg (2)			
Virgin Islands (1)	South Port (2)			
Virginia (12)	Sparta (3)			
Washington (7)	Stevens Point (6)			
West Virginia (6)	Strum (1)			
Wisconsin (192)	Superior (2)			
Wyoming (18)	Three Lakes (1)			

Figure 12. Screen capture from NewspaperArchive.com demonstrating the underlying breadcrumb trail navigation for browsing by location.

A couple of major proprietary newspaper databases have recently added a map browsing option, but these maps have limited functionality compared to other typical web-maps. They also lack geographic detail beyond state or county boundaries, or are limited to contemporary (Google) basemaps that likely would not provide adequate historical context for researchers.

The National Digital Newspaper Program (NDNP)

It is worth highlighting a couple of the major sources of newspaper data that are available to researchers, and note one interesting project that has already used this data for historical research:

The US Newspaper Program (USNP) ran from 1980 to 2007. This joint project of the Library of Congress and the National Endowment for the Humanities sought to gather bibliographic information (or metadata) for any and all newspapers nationwide (–, 1987). The result is a searchable online directory of titles that currently lists over 153,000 newspapers with links to at least 900,000 individual library holdings records (LOC, 2013).

The work of the USNP laid the groundwork for the current National Digital Newspaper Project, which has so far funded digitization of a *select number* of titles from across 39 states (Sweeney & Hawkins, 2007). Through this program, libraries and other organizations may obtain grants to fund the digitization of up to 100,000 pages from their collections. This program only includes content published between the years 1836 and 1922, though, and the resulting database of about 8.5 million pages from 1,500 newspaper titles may be only about one-percent of the total content published to date.

Given the quantity of newspaper content left to be digitized, and a political climate in Washington that seems unwilling or unable to increase financial support for this kind of work, it is unlikely that the digitization of American newspapers will be completed anytime soon. Still, the *Chronicling America* website provides a glimpse into the potential of these ongoing efforts, and serves as an important discovery and access tool for librarians and researchers.

OCLC's WorldCat

If the NDNP provides a means of *discovery*, OCLC's WorldCat search tool is probably the biggest source for gaining *access*. OCLC was founded as a cooperative organization back in the 1970s for the purpose of sharing MARC records between libraries. MARC (or machine-readable catalog) records provide libraries with standardized, searchable metadata for books, magazines, journals, newspapers, and multimedia. Typical metadata in a MARC record includes title, author, date of publication, place of publication, and so on. When a user searches a local library catalog, they are searching through that library's collection of MARC records for the titles it owns.

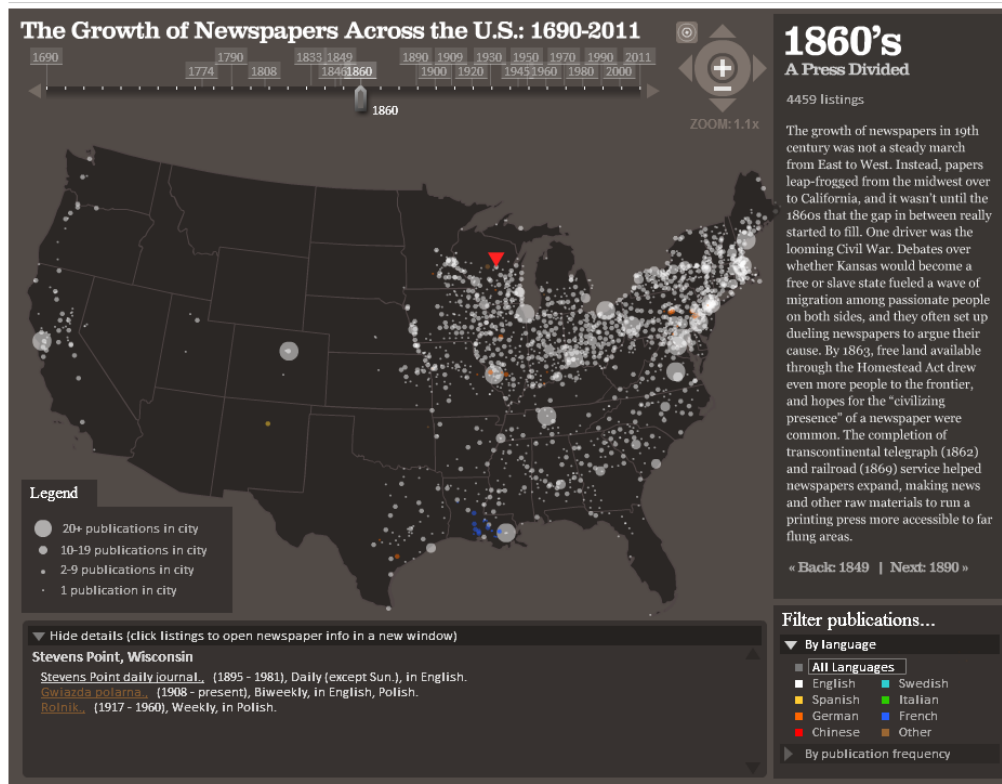
In addition to MARC data, a library may add other information to a local catalog that describes what it has available, such as how many copies the libraries has and where in the library they are located. This holdings information is often shared with OCLC, making OCLC's database (freely searchable

through WorldCat.org) a tremendous research tool for discovering physical copies of publications at libraries worldwide. The WorldCat database is the foundation for the interlibrary loan system, allowing libraries to locate and share materials back and forth around the globe. WorldCat has also added library location information into their database, and their “Find in a Library” feature allows users to locate the nearest library holding a particular item of interest.

The OCLC database is a subscription service, though, and local holdings information is added by subscribing libraries voluntarily. Smaller libraries with limited budgets or limited staff may not be able to afford to participate or may choose not to update OCLC with local holdings information. This is often true at school and corporate libraries, as well as local history museums and archives. Still, WorldCat provides access to over two-billion holdings records located at over 10,000 libraries worldwide, making it likely to be the most complete listing on the planet. The Library of Congress’ newspaper directory mentioned earlier actually links out to OCLC’s database for holdings information.

Stanford’s Rural West Initiative: Journalism’s Voyage West

Researchers at Stanford’s *Rural West Initiative* used the US Newspaper Directory to create a data visualization entitled “Journalism’s Voyage West” (Chang et al, 2012). Their website allows users to see the locations of papers published during a given year. The size of the point locations on the map indicates the number of papers published, and clicking on any one point displays titles and dates below the map.



Data sources: [Chronicling America collection](#), Library of Congress, Urban population estimates from NASA
 Data visualization by the [Rural West Initiative](#), Bill Lane Center for the American West, Stanford University
 Credits: Dan Chang, Krissy Clark, Yunkai Ge, Geoff McGhee, Yinfeng Qin and Jason Wang



Figure 13. Screen capture of Stanford's newspaper data visualization, entitled *Journalism's Voyage West*.

The project provides spatial context by including contemporary state boundaries, and the map offers typical web map options such as zooming and panning. The project creators note that the data set that they used to create this visualization is incomplete and that in many instances the actual dates of publication had to be estimated from existing data. They also noted that adding titles that not available in the original data set is difficult and they have added a Google form for anyone interested in adding a title or correcting what information they have.

Since the only geographic details the Stanford project provides are generalized contemporary state boundaries, a user's ability to pinpoint locations of interest or to place this information into historical context is limited. And since the data relies on OCLC for holdings information, its utility for facilitating access to local collections is also somewhat limited.

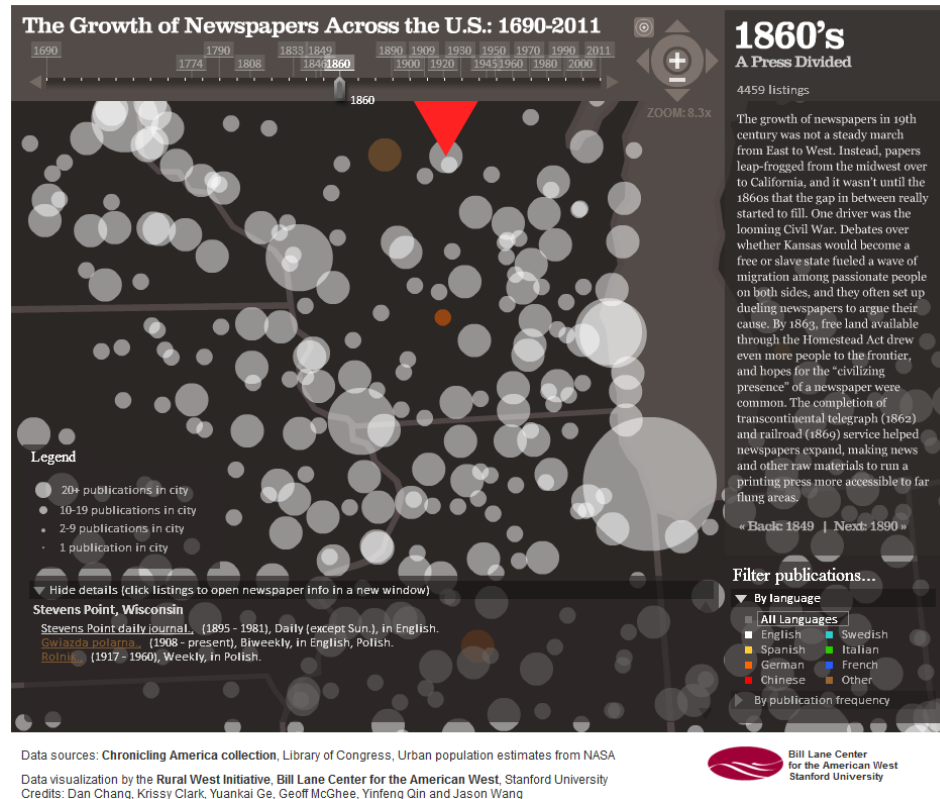


Figure 14. Screen capture of a detail of the Stanford visualization project, including southern Wisconsin.

Still, the Stanford site has already proven useful for research purposes well beyond the original intent of its creators as at least one genealogy website has already written about the its potential for use in researching family histories.

Capstone project prototypes

The goal of this capstone project, again, is to work toward a web service that would facilitate discovery and access to historical newspaper content. The original plan was to use the process roughly outlined in Appendix A. The concept of an interactive historical newspaper web service has been demonstrated by the Stanford journalism project. But to test the feasibility of a more robust web service a number of prototypes were sketched out and two were developed. Wireframing and prototyping are

standard practices in GIS project development and allow preliminary design work, testing of data organization and representation, along with other GIS planning (Tomlinson, 2013). Each of the prototypes provides a slightly different way of representing the data and in a different visual context.

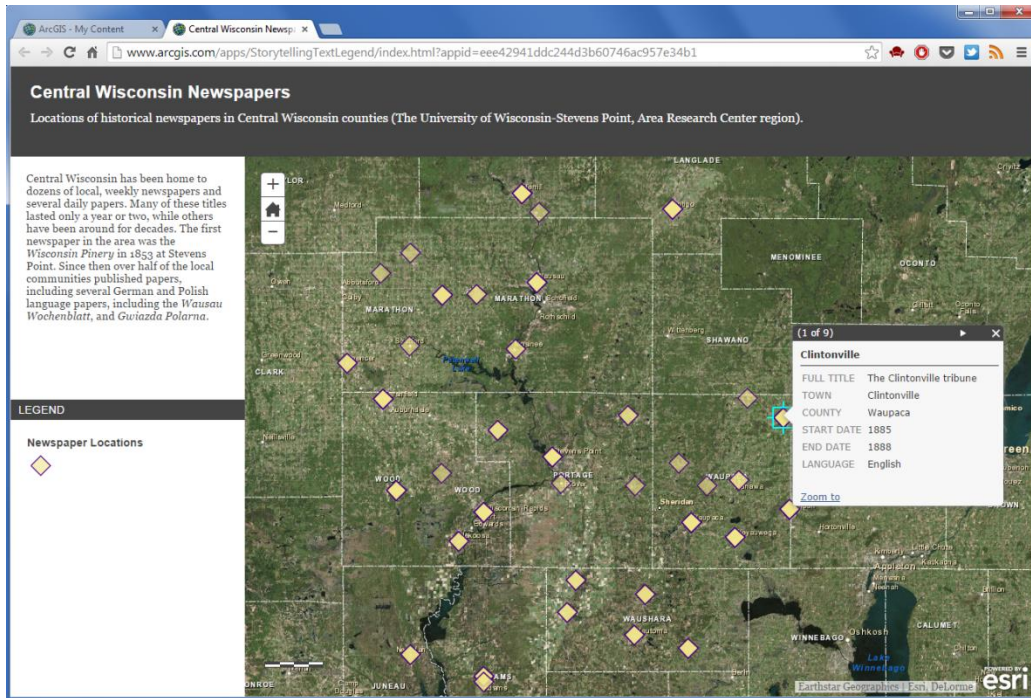


Figure 15. Prototype of a web service for locating historical newspaper publications.

The figure above is a screen capture of one of the prototypes. The spatial and attribute data for this prototype were integrated using ESRI’s ArcGIS 10.0 software package and uploaded to one of their Story Maps applications. Story Maps are a series of free apps available for licensed ArcGIS users to visualize, host, and share their data in any one of a series of pre-defined web map templates. Story Maps may be hosted on ArcGIS.com, or may be downloaded and further customized for hosting on local servers. The second option allows users a bit more flexibility for editing code in order to enable further functionality.

The header and text box in the prototype above has been edited a bit to include sample explanatory information that is intended to provide historical context for users. The point locations on the map are approximate. These are not address-level locations of newspaper publishers, but rather the locations of the populated places where those papers were published. Many historical newspapers predate the advent of the U.S. postal system, and so address level data is typically not available for use in geocoding locations. In fact, some of the oldest newspapers simply noted their office locations as something like, “Office -- Main Street, over Lee’s Jewelry store” (*Wisconsin Lumberman*, 1863). Getting coordinates for the exact locations of the original publishers would be time-consuming and well beyond the scope of this project.

The stock basemap used in this prototype is composed of aerial imagery, which provides a sense of the natural landscape of Wisconsin. Lake Winnebago is clearly visible in the lower right, and the Wisconsin River and its various flowages are visible in the left third of the screen. This imagery may be useful, depending on a user’s familiarity with the landscape. But the aerial images are visually busy, and appearance of the map is cluttered, which may distract from the map’s intended purpose.

The spatial data (locations of populated places) included in the prototype were derived from geographic coordinates downloaded from the U.S. Geographic Name Information System (GNIS). Data from the GNIS system is convenient in that it is freely available, fairly complete, authoritative, and most importantly, tends to include both current and historical place names as shown in the figure above (GNIS, 2014).

USGS
Geographic Names Information System (GNIS)

Stop! Do not bookmark or copy/paste this URL before

Feature Detail Report for: Clintonville

ID: 1563198
 Name: Clintonville
 Class: [Populated Place \(Definitions\)](#)
 History: Incorporated in 1879.
 U.S. Geological Survey. Geographic Names Phase I data compilation (1976-1981). 31.Dec.1981. Primarily from U.S. Geological Survey 1:24,000-scale topographic maps (or 1:25K, Puerto Rico 1:20K) and from U.S. Board on Geographic Names files. In some instances, from 1:62,500 scale or 1:250,000 scale maps.

Entry Date: 29-Aug-1980
 *Elevation: 817/249

*Elevations in feet/meters from the [National Elevation Dataset](#)

Federal Codes

Census Code	Census Class Code	GSA Code
15725	P1	1040

Class Code Description: **Populated Place that is also an incorporated place with the same name and the same Census Code.**

Variant Names

Variant Name	Citation
Pigeon	Citation

Counties

Sequence	County	Code	State	Code	Country
1	Waupaca	135	Wisconsin	55	US

Coordinates (One point per USGS topographic map containing the feature, NAD83)

Sequence	Latitude(DEC)	Longitude(DEC)	Latitude(DMS)	Longitude(DMS)	Map Name
1	44.6205349	-88.7623228	443714N	0884544W	Clintonville South
2	44.6180349	-88.7498224	443705N	0884459W	Bear Creek
3	44.6288681	-88.7617675	443744N	0884542W	Clintonville North
4	44.6291460	-88.7489893	443745N	0884456W	Embarass

Figure 16. Screen capture of populated place data from the US Geographic Name Information System (GNIS) for Clintonville, WI. Note the variant place name, "Pigeon."

The newspaper titles and other attributes that were linked to the populated places in this prototype were downloaded and parsed from the Library of Congress online newspaper directory, part of the *Chronicling America* digital newspaper project (LOC, 2014). Unfortunately, this data is not easy to extract as a batch, is quite often incomplete, and the information provided for each title does not always match that found in other sources or what is found on the newspapers themselves. But compared with other options, it provided a fairly convenient source for gathering a lot of titles in a relatively short period of time.

The prototype shown above is somewhat interactive in that it offers familiar contemporary web mapping options such as panning and zooming. Clicking on one of the point locations on the map opens a small pop-up window with basic attribute information about each title published at that populated place. The pop-up window allows a user to scroll through each title at that location, and to find title, date, and primary language of publication for each one. Also included is the name of the town and the county where that paper was published. This last bit of information is included because place names have often changed over time. Populated places without published newspapers were not included in this prototype.

The point locations were symbolized with diamond shapes, which provide a user with no real clues as to the purpose of the points. The number of titles at each location is indicated using transparency. The more titles published at a given location, the more solid a point appears. This is not an ideal way to represent the data, and other options, such as the size of the points, may be more appropriate for conveying information to the user.

The legend in this prototype includes only one bit of information, that diamond shapes indicate locations of newspapers. No information is provided about county boundaries or other symbology on the map, and the user is left to interpret this information on their own. From a cartographic perspective, this is a somewhat poor design for a discovery interface, but it provides a framework for discussion of the issues noted above: how should this data be represented visually in order to enable a user to understand the purpose and meaning of the map?

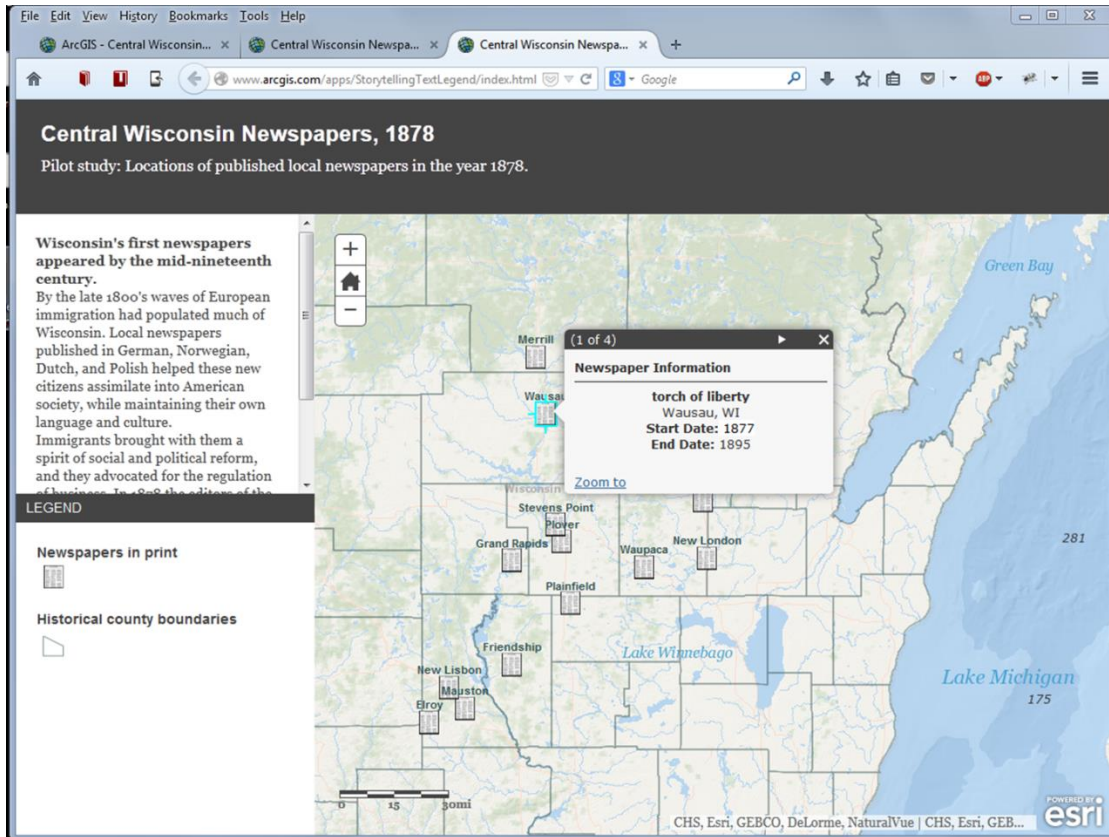


Figure 17. An earlier prototype of a web mapping service, including only location with papers published in 1878.

For the sake of comparison, the figure above is an earlier prototype of the same concept. This version shows just data from towns with papers that were in print for the year 1878. The process used in developing this prototype is more fully explained in Appendix B. The intent was to develop a series of tabs with papers available during specific years in order to examine some ways of displaying change over time.

Again, the development of multiple prototypes allows for comparing options for symbolizing and representing various aspects of the data, to view different backgrounds (or *basemaps*) that might provide geographic context, for usability testing of different interactive elements, and to help focus

feedback from colleagues and library users about the attributes and other options that may be most useful for researchers.

In this later version the stock basemap is a vector image, almost void of geographic detail. This is actually a stock basemap of the world's oceans. Stock basemaps available for web services tend to include so much visual detail as to distract from the original purpose of the map, or tend to include only current features, which may not be relevant for historical periods. The hydrographic and landcover detail on the land masses in this instance provides some degree of visual context, while providing a relatively clean canvas over which to add other layers of information. The ability to add layers of data is one of the advantages of a GIS, as compared with static map images.

In this case, county boundaries from 1878 were downloaded from the Newberry Library's Atlas of Historical County Boundaries and added. This provides users with relevant historical context, compared with something like a contemporary Google or Open Street Map. This prototype also includes a different set of point symbols that were designed to resemble a newspaper publication. This design choice may again help users understand the purpose of the map.

The two prototypes presented above are admittedly rudimentary and do not take advantage of the many capabilities of contemporary Geographic Information Systems. But they serve to help identify and work through some of the unique challenges involved in such a project.

More robust geographic discovery tools are certainly within the capabilities of contemporary GIS systems. A seemingly endless variety of web mapping services are created every day to help researchers visualize, and analyze spatially-enabled data in new ways, from census and economic data, to crime reports, to election results, and so forth.

In addition to the capabilities contemporary GIS, many library research databases and catalogs offer Application Program Interfaces (APIs) that would allow for the kinds of dynamic data access necessary for a truly interactive web map. True spatial and geographic discovery tools would be possible if the databases include consistent spatial information. This does not seem to be the case.

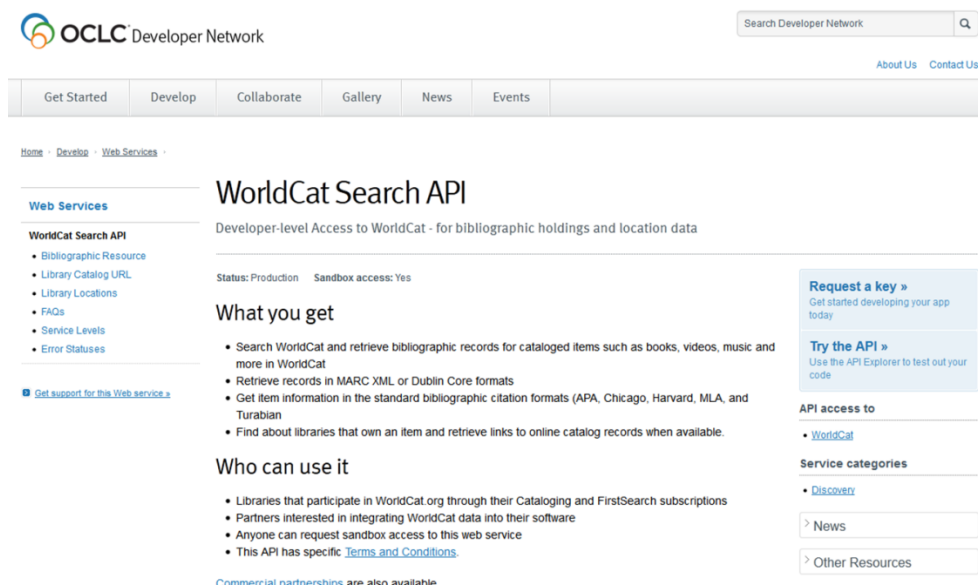


Figure 18. Screen capture of the information page for WorldCat's API.

The lack of refined geographic data in library resources persists in spite of the fact that there is popular demand for online tools that would help enable the discovery and access of historical newspaper content. Genealogy, for example, has become a major industry and the popular family history research site Ancestry.com reported over 2.1 million users during the second quarter of 2014.

But building functional, interactive web maps (or any other kind of GIS) requires mathematically definable location data (such as geographic coordinates) to join with attribute (tabular) data such as

names, dates, or population figures. It helps tremendously if the data can be linked together through a common key. And it helps if the location and attribute data is both neat and complete, which seems quite rare for historical newspapers.

Project Challenges

The challenges involved in historical GIS projects are numerous, and well-documented; I've already mentioned the lack of address-level location information for geocoding above. Finding datasets for any historical topic is difficult at best, and in most cases the data that does exist was derived from print maps that vary in both accuracy and completeness. Some topics, locations, or time periods may not even be represented in physical maps, and the landscape obviously changes over time, as does the human environment. I will highlight just a couple of challenges more specific to this pilot project, for example:

Definition

The problems start with defining what a “newspaper” actually is, given the existence of so many variations including monthly supplements, children’s newspapers, alternative press publications, and variations in format (4-pages, 8-pages, etc.). One possible definition is anything that includes a “masthead” with editor and publisher information. The definition used in this study was essentially any title that was listed in the Library of Congress directory.

Uniqueness

Then identifying unique or original “titles” requires systematically examining a variety of sources in multiple locations. Piecing together the timeline of a single “title” can be difficult, as demonstrated in

this slide, illustrating what has been called the “genealogy” of newspapers. I’ll attempt to piece together one such genealogy below:

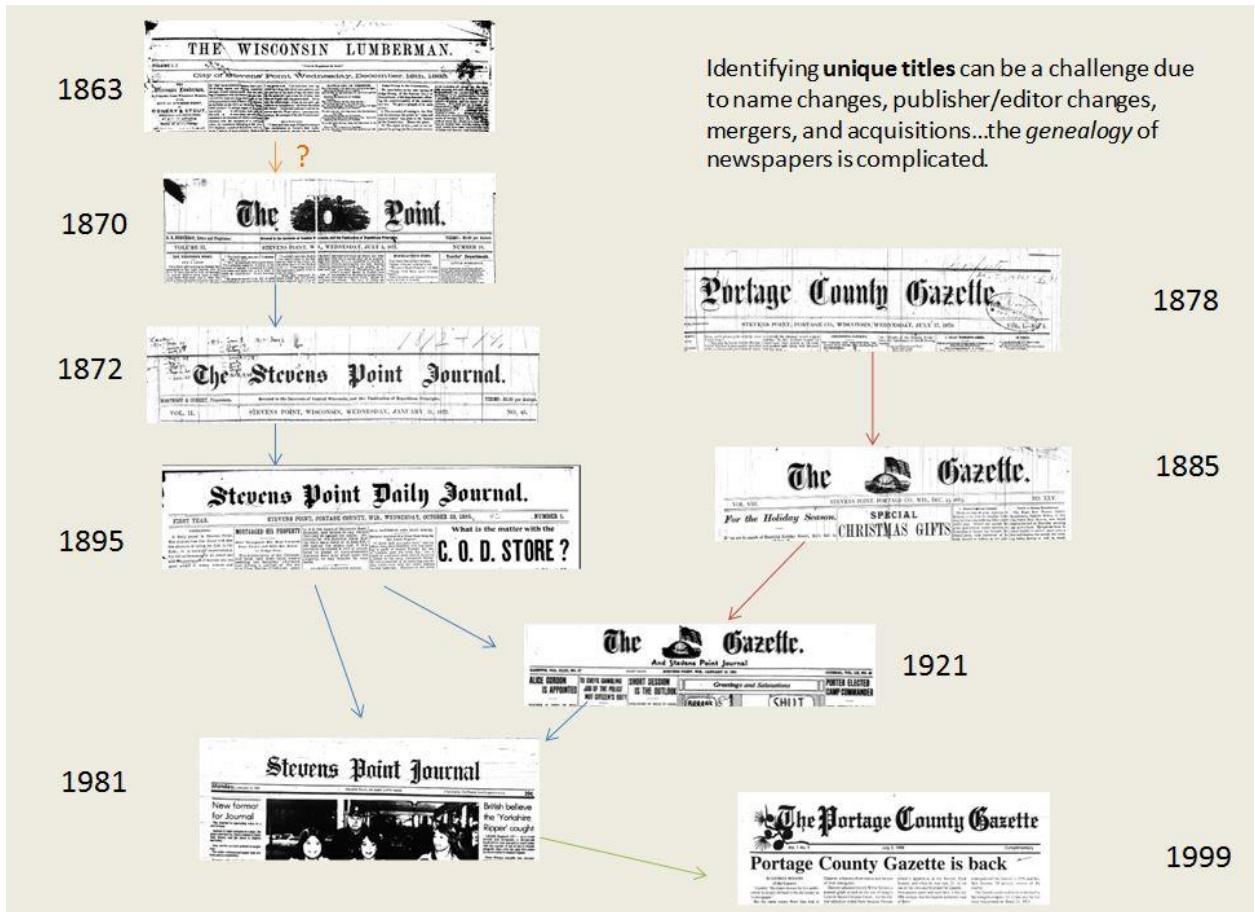


Figure 19. The complicated “genealogy” of two closely related newspapers, the Stevens Point Journal and the Portage County Gazette.

One of the first papers in Stevens Point was a weekly named *Wisconsin Lumberman* which first appeared in 1863. This paper appears to have closed a few years later, only to be reopened as *the Point* in 1870, and then renamed the *Stevens Point Journal* in 1872. *The Stevens Point Journal* became a daily and was subsequently renamed *Stevens Point Daily Journal* in 1895.

Meanwhile, a rival weekly *Portage County Gazette* started in 1878, only to be shortened to *The Gazette* in 1885. The owners of the Stevens Point Daily Journal appear to have acquired *The Gazette* in

1921 and renamed it *The Gazette and Stevens Point Journal*. This separate weekly continued just a couple of years until the *Gazette* name was dropped altogether sometime during 1923, assumedly merged in its entirety into the *Stevens Point Daily Journal*. The *Daily* changed names again in 1981 back to the *Stevens Point Journal*. So once again, in 1999, former owners and employees of the downsized *Stevens Point Journal* staff started a rival, separate weekly named *The Portage County Gazette*.

This example is fairly easy to follow but there are others far more complex. Sorting out and identifying unique titles, related titles, continuing, or preceding titles can be a complex and time-consuming process, and makes cross-referencing related data a challenge. The complexity is then compounded by changes in editor or publisher that may not have been represented by changes in a newspaper's name.

Conflicting Sources

The genealogy above was derived from multiple sources, both in print and online. These sources all too often disagree on even the most basic information such as names, dates, and other important details. Library catalogs, for example, have traditionally dropped initial articles (e.g., "The") while print publications vary in their inclusion. But there is at least one instance in this study where a newspaper chose to drop the initial article themselves as part of an official name change.

Depending on how you define a newspaper, or a unique title, there have been as many as thirty-one papers published in the City of Stevens Point (which is just one of over one-hundred communities in the pilot study area.) To examine the consistency of these sources a table was prepared using title and data information from multiple sources in order to allow cross-comparison.

Comparing Sources for Newspaper Titles				
#	WHS book by Oehlerts (covers 1833-1957) pub. 1957	LOC Newspaper Directory	OCLC WorldCat	PCHS/Book by Perret
1	Central Wisconsin Herald (1935-1938)	Central Wisconsin herald (1935-1938)	Central Wisconsin herald. (1935-?)	Central Wisconsin Herald (1935-1939)
2	Gazette (1878-1923)	The gazette (1885-1921)	The gazette. (1885-1921)	(Noted as Portage County Gazette?)
3	Gwiązda Polarna (1908-date)	Gwiązda polarna (1908-current)	Gwiązda polarna.	Gwiązda Polarna (1908)
4	Point (see Stevens Point Journal)(1870-1871)	The point (1870-187?)	The point.	(Noted as Stevens Point Journal?) (1870)
5	Portage County Gazette (see Gazette)(1878-1885)	Portage County gazette (1878-1885)	Portage County gazette. (1878-1885)	Portage County Gazette (1878)
6	Rolnik (1892-date)	Rolnik (187?-1960)	Rolnik (1891?-1960)	Rolnik (PLOWMAN) (1891)
7	Sensation (1873-1874?)	x	The Sensation	Sensation (1873)
8	Stevens Point Daily Journal (1895-date)	Stevens Point daily journal (1895-1981)	Stevens Point daily journal. (1895-1981)	Stevens Point Daily Journal (1895)
9	Stevens Point Democrat (1880-1885)	Stevens Point democrat (1880-1885)	Stevens Point democrat. (1880-?)	Stevens Point Democrat (1880-1885)
10	Stevens Point Journal (1870-1921)	The Stevens Point journal (1872-1921)	Stevens Point journal. (1872-1921)	Stevens Point Journal (1872)
11	Stevens Point Post (1892-1899?)	x	x	Stevens Point Post (German) (1892-1899)
12	Wisconsin Eagle (1884-1885)	Wisconsin eagle (1884-1885)	Wisconsin eagle. (1884-1885)	Wisconsin Eagle (1884-1885)
13	Wisconsin Lumberman (1863-1868)	The Wisconsin lumberman (1863-1868)	The Wisconsin lumberman. (1863-?)	Wisconsin Lumberman (1863)
14	Wisconsin Pinery (1853-1893)	The Wisconsin pinery (1853-1893)	The Wisconsin pinery. (1853-?)	Wisconsin Pinery (1853)
15	Wisconsin State Press (1894-1896?)	x	x	Wisconsin State Press (1894)
16	Wisconsin States Rights (1859-1860, 1861-1862)	Wisconsin state rights (1859-1862)	Wisconsin state rights.	Wisconsin State Rights (1859, 1861)
17	x	The Central city herald (1935-1935)	The Central city herald. (1935-1935)	(Noted as Central Wisconsin Herald)
18	x	The Gazette and Stevens Point journal (1921-197?)	The Gazette and Stevens Point journal. (1921-?)	(Notes consolidation?) (1919)
19	x	Horyzonty (1989-current)	Horyzonty.	x
20	x	Stevens Point journal (1981-current)	The Stevens Point journal. (1981-date)	(Noted as continuation of Stevens Point Daily Journal) ()
21	x	x	The Portage County gazette. (1999-date)	Portage County Gazette (1999-date)
22	x	x	Whipping post	x
23	x	x	Counterpoint	x
24	x	x	GP Light : monthly supplement to the Gwiązda Polarna	x
25	x	x	x	Hungerford's Real Estate Journal (1879-1886)
26	x	x	x	Stevens Point Press
27	x	x	x	Niedziela (Sunday) (1872-1874)
28	x	x	x	Jaskolka (Swallow) (1930-1935)
29	x	(noted with Gwiązda Polarna)	(noted with Gwiązda Polarna)	Stoneczko (A Little Sun) (1938-1940)
30				aka. Pinery
31				aka. Weekly Pinery

Figure 20. A comparison of sources of newspaper title and date information. Different sources may provide different titles, and sometimes conflicting information on the same title.

The figure above provides information derived from five sources, including Oehlerts' directory, the Library of Congress' online Newspaper Directory, OCLC's WorldCat, the local Portage County Historical Society's website, and a print publication on the various holdings at the State Historical Society in Madison. Note that dates here are reduced to just the year of publication, as months and days are simply unavailable in some sources. There has proven to be quite some variation between sources, and collecting and comparing data from multiple sources may be a necessity.

Probably the most intriguing title is the one in row 7, named *Sensation*. This title, which is noted with uncertain dates by my first source, is not listed in the second, listed without dates but with the initial article "The Sensation" in the third source, listed with just one year of publication in the fourth source, and absent from my final source. Even the smallest variation raises questions about the accuracy or completeness of a particular source.

By one description the *Sensation* title appears to have been satirical publication, something like a nineteenth century version of *The Onion*. So again, deciding whether or not to include this in a newspaper database would depend on your definition of "newspaper."

Selecting Attribute Data

The sources mentioned above were created by different organizations, with different purposes, and hence they may provide different kinds of information about the same title. In the figure below I've identified approximately twenty-six possible attributes from across four sources. Any of these might be used to populate a newspaper database, but some of these (such as title) are more likely to be useful than others (such as editor). A best guess, based on personal experience, is that title, variant titles, and dates of publications would most likely be useful in the project context, along with the name(s) of the community where it was published, both at the time of publication, before, or since.

Attribute data available by source				
	Oehlerts	Danky & Hady	LOC	WorldCat
1	Title	Title	Title	Title (minus initial article)
2	Variant titles	Alternate titles	Alternative titles	Other titles
3	Dates of publication	Dates	Dates	Dates
4	Frequency (d, w, m)	Frequency	Frequency	x
5	Editors/publisher	Publisher	Publisher (or editor?)	Publisher (or editor?)
6	WHS or other holdings at time	WHS holdings	Link to library holding:	Link to library holdings
7	Dates of holdings	WHS dates of holdings	x	x
8	County	County	County	x
9	City	x	City	City
10	Language	x	Language	Language
11	x	DLC#	x	x
12	State (implied)	State (implied)	State	State
13	x	x	Geographic coverage*	x
14	x	x	LCCN#	x
15	x	x	OCLC#	OCLC#
16	x	x	Description	Description
17	x	x	Notes	Notes
18	x	x	Subjects	Subjects (e.g., "Stevens Point (Wis.) -- Newspapers.")
19	x	x	x	Genre
20	x	x	x	Material type
21	x	x	x	Document type
22	x	x	x	Related subjects
23	x	x	x	Reviews (if any)
24	x	x	x	Tags (if any)
25	x	x	x	Similar items (if any)
26	x	x	x	Linked Data*

Figure 21. Possible attributes for populating a newspaper database.

Verification

The variations in the kinds of data provided in these sources, as well as gaps, inconsistent, inaccurate, or incomplete data makes verification of attributes using the original source material often necessary. But verification of data is time-consuming and has proven well beyond the scope of this capstone project. A possible method for collecting and verifying larger amounts of newspaper data (while minimizing time spent spinning through reels of microfilm at the local library) might be as follows:

1. Download and create a table of attributes using the Library of Congress directory information (title, dates, location, etc.) for a given community.
2. Review OCLC WorldCat, a trusted print index, or another reputable source to compare the same title information.
3. If the information in both (or all) sources agrees, consider it correct and move on to the next title in the community. (This assumes that the data for each source did not itself come from the same source).
4. If there is a conflict between these data, proceed to track down and verify the information from the original source material (i.e. the first issue of a new title, if possible)
5. When one community is complete, move on to the next...

There are many potential sources that may be used for collecting and verifying data. One resource is available online via the Library of Congress' digital collections, known as "Rowell's and Ayers' newspaper directories." Others may include the guides and handouts created by state historical societies or other organizations with more familiarity with their local publications.

Another possible option is to allow the users of the system, or other volunteers, to find and correct errors themselves. The concept of Volunteered Geographic Information has been used in several projects where resources are lacking. With a bit of promotion, "Volunteered Librarian Information" may

even become a thing, allowing library and archival staff from around the area to verify newspaper data and holdings information for their own communities. The current state of library budgets makes this unlikely at present.

Incorporating Temporal Data

The figure below provides a visual timeline for Stevens Point papers during the second half of the nineteenth century. This was created in order to visualize temporal data, and to explore the possibility of including a time-select option (such as by year) in a discovery interface. Again, the dates provided may be inexact due to gaps in the data, but it appears that only three of these titles lasted 11 or more years. This volatility may help explain why there are so many titles in the Library of Congress directory.

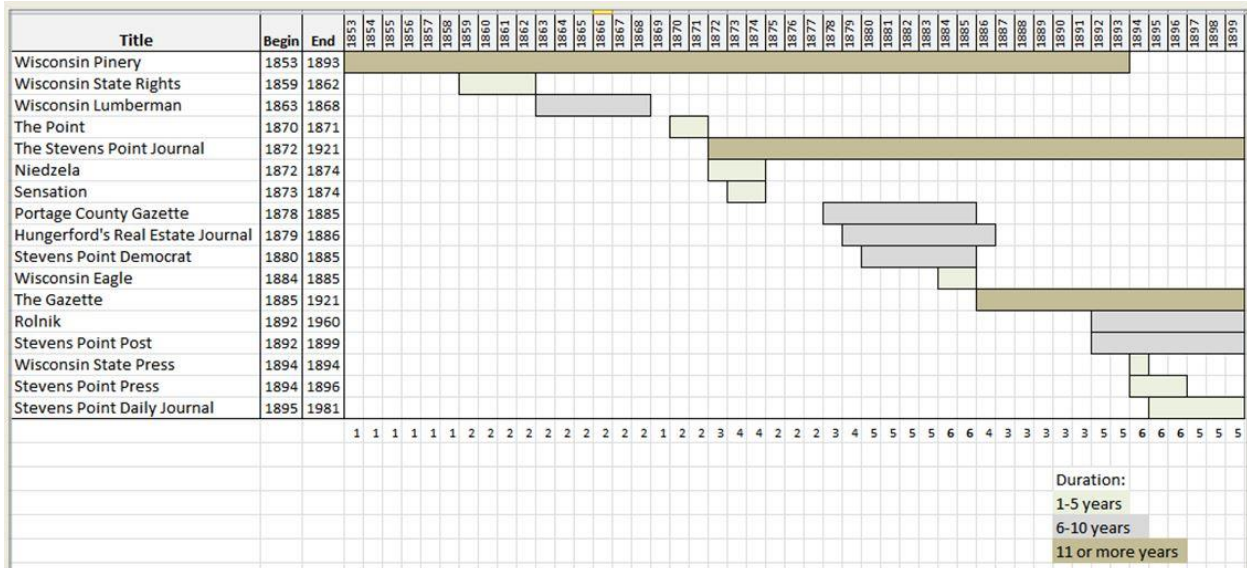


Figure 22. Late-19th Century timeline of local newspapers, Stevens Point, WI, USA.

In creating this timeline, certain assumptions were made about what would constitute a “unique” title. For example, here I count a name change as constituting a separate title (though some of these, using the genealogy presented above, could arguably be merged into one continuous lineage). Using this interpretation, it appears that as many as six papers were published in the City of Stevens Point at a given time, with no fewer than two in print after the 1850s.

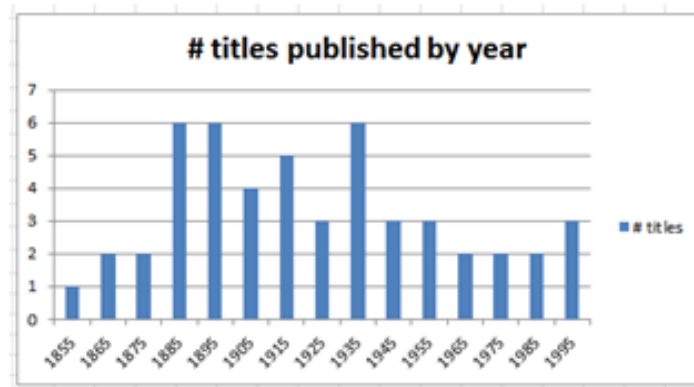


Figure 23. The number of titles by year published in Stevens Point (in ten-year increments 1855 to 1995).

Determining the “Geography” of a newspaper?

Assuming that a complete and accurate record of one newspaper’s attribute data has been compiled, the decision to tie a specific newspaper to a specific geography is debatable. Several papers in the pilot study area may demonstrate the complexities of assigning concrete locations to newspapers, for example:

- The *Stevens Point Journal* is named for the City of Stevens Point, reports events within city limits but also covers surrounding towns, rural areas, state-wide, national, and even international news. Any and all of these geographies could be assigned to this title.

- The *Portage County Gazette* is also published in the City of Stevens Point, but the title implies a journalistic focus on issues relating to the county that contains the city. Again, multiple levels of geography may be appropriate.
- The *Central Wisconsin Herald* was published for a short time in Stevens Point, but the title obviously implies a broader and more amorphous geography. There is no clear definition of the term “Central Wisconsin” even though this is a commonly used phrase in the pilot study area.
- *Gwiazda Polarna* is a Polish-language paper published for decades in the City of Stevens Point, but this title appears to have been intended to appeal to a broader, national, Polish-American audience.
- The title *Wisconsin States Rights* may have been an early attempt at a statewide newspaper (like the more contemporary *Wisconsin State Journal*). A statewide geography is implied, and *Wisconsin States Rights* was published at different times in both Stevens Point and Monroe, WI.
- Finally, the Wausau area German-language paper, *Wausau Wochenblatt*, later known as the *Wisconsin Wochenblatt*, was published in both Wausau, WI and Winona, MN.

The decision to use publisher location in the prototypes above is certainly not without its problems, but it allows for a convenient method of geolocation. The option to assign multiple levels of geographies to individual newspaper titles may be warranted, but is beyond the scope of this project.

Local Newspaper Holdings

So far this discussion has focused on collecting the attribute data that would enable *discovery*. With the discovery pieces in place, the *access* part of the project provides its own set of challenges. For example, In the State of Wisconsin there are currently twenty-six University of Wisconsin System campuses, including several larger campuses with multiple libraries, along with many private universities and technical colleges , almost four-hundred public libraries, eighty-two public branch libraries, and

dozens of large and small state, county, and local historical societies. Each of these institutions may potentially provide access to any number of newspaper holdings. While identifying these collections and collecting basic information about the institutions (physical address, website, etc.) may not be particularly challenging, information about the holdings available at each may not be easily accessible, requiring a site visit or some personal contact.

For institutions with holdings information available in OCLC, there are additional concerns. A record for a single title may be linked to multiple holdings, each with varying periods of coverage, and those may be in print, on microfilm, or online in a digital format. Connecting users to these holding may be accomplished in part by including common keys in an attribute table (e.g., OCLC accession numbers) found in WorldCat or in online databases.

Examination of various options has revealed a number of problems with this route. Some records in OCLC, for example, provide incomplete or simply inaccurate holdings statements, that do not actually reflect what is currently available. Other records may provide inaccurate holdings data for a given title, or may not correctly identify separate holdings for a title that has changed.

Thousands of libraries outside of the state have also collected papers from this pilot area, as demonstrated below:

of Congress' *Chronicling America* project, mentioned above, has not digitized any Wisconsin newspapers to date.



Stevens Point, WI
Dates: 1892-1960
Frequency: Weekly
Language: Polish
Variant Titles:
Plowman
Locations:
UWSP
PCPL
WHIS

Figure 25. Pop-up information window, including a nameplate, location, dates of publication, frequency, language, variant title, and abbreviated locations of library holdings.

Putting all of this holdings data together would again involve a lot of time spent on data collection and verification. Incorporating this information into a web service might be accomplished more simply by linking out to any organization that appears to hold a title, without including the holdings information itself. It would be up to the user to explore that location's collections in depth in order to determine what dates are actually available. This is obviously not ideal, but may be the only practical solution until the bulk of digitization work nationwide is completed.

Prototype Evaluation

To begin evaluating the merit of a historical newspaper discovery resource from a user standpoint several staff at the university archives were invited to view the prototypes above and were asked for their feedback. After a brief demonstration and introduction to the project, the archives staff was asked a number of questions. Their comments are summarized briefly below:

- What are your first impressions, would this kind of tool be useful? *Yes, definitely. Genealogists would love this.*
- Who are the most likely potential users? *Local community members and library visitors, particularly those with an interest in genealogy. Instructors on campus, and other librarians and archivists in the area. Also, History students in a course with an assignment to research the history of a particular town. Local historians as well.*
- What are some potential research areas/applications? *Family history, obviously. Also environmental history. Health research. Anyone studying issues that involve the growth and transformation of human population centers.*
- What attributes are most likely to be useful? *Location of the newspaper (name of town/place); Title of newspaper; dates of publication; possibly editor/publisher, but that is not a priority.*
- Same question as above, but with the collected table of possible attributes on the screen. *Also, frequency of publication; library holdings (via WorldCat); County (which has changed with time); Language of publication.*

Based on this preliminary evaluation and after conversations with several librarians and other expert researchers it appears there is indeed some merit to the concept. Persons familiar with the complexities of historical newspaper research found the idea of a geographic discovery service

intriguing, and they imagined many possible uses, both in an academic and research library setting or just in terms of popular use on the web for genealogy, family history, etc. While I was pleased to find that some of my assumptions about attribute selection and likely users were confirmed by my colleagues, I was warned by an experienced historian about "going down the rabbit hole" trying to sort out the mess of historical newspapers.

Conclusions

In summary, there are several major sources of data with potential utility for developing a geospatially enabled discovery tool for the thousands of historical newspapers published and collected at locations across in the United States. But attribute data in these sources is often flawed or incomplete and requires time and effort to pull together into a consistent framework for presentation in a web mapping service. In addition, newspapers can represent various, shifting geographies. There are many issues yet to solve in order to move the concept of a geospatially-enabled, historical newspaper discovery service forward.

There are a number of ways this idea could be managed at a broader geographic (statewide, national) level. One option would be for a major organization such as the Library of Congress to use their data as is, and develop a single map interface, noting the issues inherent and excluding the most problematic data altogether. The Library of Congress could also initiate another nationwide data collection and verification project, which seems unlikely.

Another option would be for an organization with the available resources to develop a system that allows users themselves to correct problems with the data as they are discovered. This concept is generally referred to as Volunteered Geographic Information (VGI) and is the method currently being

used by the U.S. Geological Survey in the National Map project. This method probably holds the most potential for developing and populating a comprehensive, nationwide or even international web service.

A more likely scenario is for smaller, regional, or possibly statewide projects to be completed by organizations like the Wisconsin Historical Society or the UW System Libraries. This would be beneficial in that a degree of local knowledge is quite useful in sorting out the numerous problems involving names, dates, locations, etc.

The plan for this particular project is to continue with collection and verification of information on titles in the pilot study area in order to build a clean and somewhat more complete dataset. That dataset will be used to continue developing prototypes that enable further study and decision-making about appropriate levels of representation, geolocation, cross-referencing, functionality, usability, and so forth. These examples will be shared with colleagues at other institutions in order to demonstrate the possibilities and to gauge interest in larger collaborative projects. It is hoped that this capstone project paper will inspire more work and collaboration on this concept.

References

- . (1987). Project & resource update: In support of scholarly research. U.S. Newspaper project rolls on. *American Libraries*. Retrieved from www.jstor.com.
- Beall, J. (2010). Geographical research and the problem of variant place names in digitized books and other full-text resources. *Library Collections, Acquisitions, and Technical Services*, 34, 74-82.
- Bonnell, J. and M. Fortin. (Eds.). (2013). *Historical GIS research in Canada*. Calgary, AB: University of Calgary Press.
- Buckland, M., Chen, A., Gey, F. Larson, R., Mostern, R. and V. Petras. (2007). Geographic search: catalogs, gazetteers, and maps. *College and Research Libraries*, 68, 376-387.
- Bailey, T. J., and J. B. M. Schick. (2009). Historical GIS: Enabling the collision of history and geography. *Social Science Computer Review*, 27(3), 291-296. Retrieved from <http://ssc.sagepub.com>
- Chang, D., Clark, K., Ge, Y., McGhee, G., Qin, Y. and J. Wang. (2012). *Data visualization: Journalism's voyage west*. Retrieved from https://www.stanford.edu/group/ruralwest/cgi-bin/drupal/visualizations/us_newspapers
- Clark, K. and G. McGhee. (2012). *Did the west make newspapers, or did newspapers make the west?* Retrieved from <http://web.stanford.edu/group/ruralwest/cgi-bin/drupal/content/rural-newspapers-history>
- Clough, P., Tang, J., Hall, M. M., and A. Warner. (2011). Linking archival data to location: a case study at the UK National Archives. *Aslib Proceedings*, 63(2/3), 127 – 147. Retrieved from <http://emeraldinsight.com/0001-253X.htm>
- Danky, J.P. and M.E. Hady (Eds). (1994). *Newspapers in the State Historical Society of Wisconsin: A bibliography with holdings. Volume 1, Wisconsin Newspapers*. Norman Ross: New York, NY.
- Gregory, I. N. and P. S. Ell. (2008). *Historical GIS: Technologies, methodologies and scholarship*. New York, NY: Cambridge University Press.
- Gregory, I. N. and A. Geddes. (2014). *Toward spatial humanities: Historical GIS and spatial history*. Bloomington, IN: Indiana University Press.
- Gregory, I. N. and R. G. Healey. (2007). Historical GIS: structuring, mapping, and analysing geographies of the past. *Progress in Human Geography*, 31, 638-653.
- Harvey, R. (1979). *The Wisconsin Newspaper Association: a history, 1853 to 1979*. Madison, WI: Wisconsin Newspaper Association.

- Holdsworth, D. W. (2003). Historical geography: New ways of imaging and seeing the past. *Progress in Human Geography*, 27(4), 486-493. doi: 10.1191/0309132503ph440pr.
- Jones, R., Zhang, W. V., Rey, B., Jhala, P., and E. Stipp. Geographic intention and modification in web search. *International Journal of Geographical Information Science*, 22, 229-246.
- Knowles, A.K. (2008). *Placing history: How maps, spatial data, and GIS are changing historical scholarship*. Redlands, CA: ESRI Press.
- Larson, R. R. (1996). Geographic information retrieval and spatial browsing, In Smith, L. and Gluck, M. (Eds.) *GIS and libraries: Patrons, maps, and spatial information*. Urbana-Champaign, IL: University of Illinois, Urbana-Champaign. Retrieved from <http://www.ideals.illinois.edu/handle/2142/416>
- Library of Congress. (2013). *Chronicling America: Historic American newspapers*. Retrieved from <http://chroniclingamerica.loc.gov/>
- Long, J. (2010). Atlas of historical county boundaries. Retrieved from <http://publications.newberry.org/ahcbp/index.html>
- Minnesota Population Center. (2011). *National Historical Geographic Information System: Version 2.0*. Minneapolis, MN: University of Minnesota. Retrieved from <http://www.NHGIS.org>
- Nesbit, R.C. (1973). *Wisconsin: A history*. Madison, WI. University of Wisconsin Press.
- OCLC Developer Network. (2014). *WorldCat Search API*. Retrieved from <http://www.oclc.org/developer/develop/web-services/worldcat-search-api.en.html>
- Oehlerts, D. (1958). *Guide to Wisconsin Newspapers, 1833-1957*. State Historical Society of Wisconsin: Madison, WI.
- Owens, J.B. (2008). Historical Studies, GIS for. In K.K. Kemp (Ed.), *Encyclopedia of geographic information science* (pp. 220-221). Los Angeles, CA: SAGE.
- Palacio, D., Cabanac, G., Sallaberry, C., and G. Hubert. (2010). On the evaluation of geographic information retrieval systems: Evaluation framework and case study. *International Journal on Digital Libraries*, 11(2):91-109. <http://link.springer.com/article/10.1007/s00799-011-0070-z>
- Perret, M.E. (1992). *Portage County, of place and time: a historical geography of Portage County, Wisconsin*. Stevens Point, WI: Cornerstone Press.
- Purves, R. and C. Jones. (2011). Geographic information retrieval. *SIGSPATIAL Special*, 3(2):2-4. Retrieved from <http://dl.acm.org/citation.cfm?id=2047297>
- Sweeney, M. and L. Hawkins. (2007). The National Digital Newspaper Program: Building on a firm foundation. *Serials Review*, 33. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0098791307000676>

- Tomlinson, R. (2013). *Thinking about GIS: Geographic Information System planning for managers*. Redlands, CA: ESRI Press.
- U.S. Geological Survey (2013), 19810501, *U.S. Geographic Names Information System (GNIS)*: U.S. Geological Survey, Reston, VA. <http://geonames.usgs.gov>
- Wisconsin Cartographers' Guild. (1998). *Wisconsin's past and present: a historical atlas*. Madison, WI: University of Wisconsin Press.
- Wisconsin Demographic Services Center. (2013). *January 1, 2013 Final Population Estimates*. Retrieved from <http://www.doa.state.wi.us/Divisions/Intergovernmental-Relations/Demographic-Services-Center/Wisconsin-Population-Housing-Estimates/>
- Wisconsin Historical Society. (2014). *Short history of Wisconsin*. Retrieved from <https://www.wisconsinhistory.org/topics/shorthistory/>
- Wisconsin Lumberman. (1863, December 16). "Wisconsin Lumberman" [Advertisement], 1(1), 5. Retrieved from <http://access.newspaperarchive.com>
- Yarasavage, N., Butterhof, R. and C. Ehrman. (2012). National Digital Newspaper Program: a case study in sharing, linking, and using data. Poster session presented at JCDL '12 12th ACM/IEEE-CS Joint Conference on Digital Libraries. Washington, DC, USA — June 10 - 14, 2012. Retrieved from http://www.loc.gov/ndnp/guidelines/docs/JCDL2012_ndnp.pdf

Appendix A –*Capstone project, proposed methodology.*

1. Creation of an attribute database of select newspaper data.
2. Joining this data to shapefiles of point locations using geographic coordinates.
3. Preparation of a series of maps from this and other historical data using ArcGIS.
4. Customizing display, symbology, legends, etc., for optimum viewing.
5. Exporting maps as shapefiles to load into ARGIS Online.
6. Publishing this map series using ESRI’s StoryMaps “tabbed viewer” application.
7. Adding explanatory text including major historical developments.
8. Customizing appearance of the Story Map application.
9. Adding user-side functionality such as mouse-over, mouse-clicks, hyperlinks, etc.

Appendix B – Capstone project, steps for first prototype development.

- Downloaded historical county boundary shapefile for Wisconsin from the Newberry Library and extracted a single year (1878). Created a map in ArcGIS, and added this data as a layer.
- Address-level data is not available for this time period, in some instances location is identified in the masthead as vaguely as “located above Johnson’s store.” I’m not sure that level of accuracy is really that meaningful for this purpose, so I decided that locating a title at the city or populated place level is the best I can do. I downloaded and extracted a csv file of current populated places for Wisconsin from the GNIS website (USGS, 2013). The GNIS does include some historical name information, which is helpful.
- Using Microsoft Excel, I extracted the populated places for my study area and eliminated any extraneous data. I selected just those places listed in LOC database, and added them as a layer using the GNIS designated X-Y coordinates in ArcMap.
- Downloaded a list of LOC Newspaper Directory titles for my selected counties. I imported the directory data, including title, dates, and location (by name of city), into a new spreadsheet. In ArcMap, I joined this data to the layer of point locations for populated places. One problem here is that multiple titles for a location creates overlapping points, which is a problem I haven’t satisfactorily resolved.
- Exported both the counties and places as shapefiles, and zipped them up for loading into ArcGIS online. Then I created a new ArcGIS map, and loaded my test data layers.
- Configured the color and size of my county boundaries, added customized icons, and added and edited a pop-up window. I also added one of ESRI’s stock basemaps for more spatial reference.
- The final step was to create a new web application using ESRI’s Story Maps tabbed browser option. I added some relevant textual information and the map was ready to publish and share.