



---

---

## *Digitizing Worcester's History*

---

---

A student-based undertaking submitted in the form  
of an Interactive Qualifying Project at  
Worcester Polytechnic Institute.

Submitted by:

Brendan Connolly    [bconnolly@wpi.edu](mailto:bconnolly@wpi.edu)  
Peter DiMaggio      [pndimaggio@wpi.edu](mailto:pndimaggio@wpi.edu)  
Daniel Grande        [dgrande@wpi.edu](mailto:dgrande@wpi.edu)

Submitted to:

Prof. Joseph Cullon, Project Advisor

Sponsoring Organization:

The Worcester Historical Museum, as represented by William Wallace

May 5, 2015

This report represents work of WPI undergraduate students submitted to the faculty as evidence of a partial degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review. For more information about the projects program at WPI, see <http://www.wpi.edu/Academics/Projects>

## *Abstract*

The Worcester Historical Museum houses a large collection of historical city directories. To access these directories, a user must visit the museum or contact its librarian. The size and condition of surviving directories make searching across them difficult and time consuming. This IQP investigated strategies to scan the directories and manage their digital contents online. Over the course of digitizing nine directories between 1850 and 1920, this IQP devised best practices for scanning and making them OCR searchable. Further the team investigated commercial and open-source data management platforms, recommending that the WHM pilot three systems. Finally, this team created three digital sandboxes with several directories loaded onto each and suggested appropriate evaluative criteria.

## **EXECUTIVE SUMMARY**

The Worcester Historical Museum (WHM) houses a large collection of city directories that chronicle the history of the inhabitants of Worcester. Currently, the only method for accessing these directories is through a visit to the museum or through contact with its librarian, Robyn Conroy. These city directories can be large in size, over 1000 pages, and, therefore, can be time consuming to search through. Therefore, the WHM museum is looking to digitize these directories and make them keyword searchable online. By having the directories available online, they will be easily accessible by any end-user, whenever and wherever it is convenient. Furthermore, a proper data management system will allow for a user to keyword search through multiple directories simultaneously and quickly find the desired result.

Creating a digital archive of Worcester city directories can be broken into three stages: scanning of the city directories, performing Optical Character Recognition (OCR) on the directories, and evaluating the best digital storage service for making them searchable online. Over the course of this project we have successfully scanned nine city directories, and performed OCR on all of them. In addition, the search for a service to make the directories available online has been narrowed to three options: ContentDM, SimpleDL, and Omeka. Each service has varying advantages and disadvantages and should be evaluated independently by the WHM to determine which one best fits their needs, capabilities, and budget.

### **Findings**

#### **Scanning of the city directories is most efficient using the cradle scanner.**

After attempts with both the bed scanner and feed scanner it was discovered that the most efficient scanning method is to use the cradle scanner. Through experience it was determined the best method to operate the cradle scanner is by employing three people. One person operates the

computer that takes and saves the images of the pages, another turns each page as it is scanned, and the third operates the cover that presses the document down before it's scanned. The feed scanner is a tempting alternative since it can automatically scan a large stack of pages quickly. However the feed scanner struggles to scan consistently, getting jammed often. We believe this to be related to the size of the page being scanned and the brittleness of the paper, as some pages scan easily while others are impossible.

**The pages scanned should be saved as a PDF file containing 30 pages.**

When scanning it is important to save often to preserve data and the PDF format is best for this project. Saving should be performed every 30 pages, as waiting any longer risks the software freezing and scans being lost. The pages should be saved as PDF files, as it is easy to perform OCR on them, and saving as TIFF files takes up a large amount of storage space. Also, it is important to name the files according to that year's directory and pages that were scanned.

**Optical Character Recognition should be performed with Adobe Acrobat XI.**

Performing OCR on PDF files is a simple process that automatically extracts the text from a PDF file and makes it serviceable. There are two types of OCR: dirty and clean. Dirty OCR is easily accomplished using Adobe Acrobat XI, while clean OCR is done professionally.

**ContentDM, Omeka, and SimpleDL should be considered for making the directories available online.**

These services all provide a simultaneous search function, metadata support, and have hosting services. ContentDM and SimpleDL are paid services, while Omeka is free (excluding the hosting services). Consequently, ContentDM and SimpleDL offer customer support, whereas Omeka users must rely on documentation or forums.

**Each service should be evaluated by the WHM using a set of variables.**

The services should be evaluated for usability, search function, relation to other digital assets, costs, security, and technical expertise.

### Recommendations

**The services should be tried by the WHM before making a decision.**

Each service has a method of trying it for free or low cost. These trials should be used by the WHM in order for them to make an informed decision on which service best fits their needs and capabilities. When trying the services, the set of variables outlined in the findings should be considered.

**The scanning process should be continued to build the collection of digitized city directories.**

Directories should continue being scanned in order to provide more information digitally to the user. Instead of going through the collection scanning for every year in order, first a directory should be scanned from every ten-year period in the collection. Once that's complete, fill in the collection with a directory for every five years of history. Continue narrowing the gap between each directory scanned until the collection is complete. This way the collection always provides a wide range of dates for the user even as it is being added to.

## *Acknowledgments*

Our group accomplishments would not have occurred if we had not received significant help from others involved. This project used many different resources across the Worcester Historical Museum and WPI. Everyone who helped during the project deserves as much credit as the team members.

Our project sponsor, the Worcester Historical Museum, gave us the guidance we need to successfully digitize some of Worcester's history. Their constant availability allowed us to pick up city directories when it was convenient for us. They also gave us great feedback on our research that led to our final presentation. It was a great experience working with the Worcester Historical Museum and we're looking forward to seeing great results in the future.

Everyone at the WPI Archives deserves credit for all the help they gave us. Without them, there would be no scanned city directories. They taught us how to use the machines and programs that delivered the PDF files of the directories. Appointments were simple to set up and we always received a response quickly.

Without our project advisor, Professor Cullon, this project would not have been successful. His continued guidance and support steered us into a positive direction and was always there to answer any questions. His advice was invaluable throughout the entire process and was a major asset towards the success of the project.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>i</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>ii</b>
Findings	ii
Recommendations	iv
<b>ACKNOWLEDGMENTS</b> .....	<b>v</b>
<b>CHAPTER 1. INTRODUCTION</b> .....	<b>1</b>
<u>1.1 Mission</u>	2
<u>1.2 Values</u>	2
<u>1.3 Problem Outline</u>	3
<b>CHAPTER 2. ENVIRONMENTAL REVIEW</b> .....	<b>5</b>
<u>2.1 Scholarly Literature</u>	5
<u>2.2 Digital Collections and Best Practices</u>	8
<u>2.3 Applications in the Digital Humanities</u>	13
<b>CHAPTER 3. METHODS</b> .....	<b>15</b>
<u>3.1 Scanning</u>	15
<u>3.2 Storage</u>	16
<u>3.3 Image Manipulation</u>	17
<u>3.4 Optical Character Recognition</u>	17
<b>CHAPTER 4. DATA MANAGEMENT SYSTEMS</b> .....	<b>19</b>
<u>4.1 Trial One- ContentDM</u>	19
<u>4.2 Trial Two - SimpleDL</u>	20
<u>4.3 Trial Three – Omeka</u>	21
<u>4.4 Beta Test of ContentDM, SimpleDL, and Omeka</u>	22
<b>CHAPTER 5. FINDINGS AND RECOMMENDATIONS</b> .....	<b>26</b>
<u>5.1 Findings</u>	26
<u>5.2 Recommendations</u>	28
<b>REFERENCES</b> .....	<b>31</b>

## CHAPTER 1: INTRODUCTION

Worcester's rich history teems with industrious people who, through hard work and dedication, drove the city from its early beginnings to an era of great prosperity and productivity. When briefly reflecting on who these people were, some obvious names come to mind, such as Percy, Salisbury, Norton, Washburn and Moen. Though these men led Worcester and played pivotal roles in its history, the city's larger population propelled the city forward. This component comprised common working men and women, small business owners, and eager immigrants. Whether it was the humble owner of a butcher shop or even just a factory worker who shopped there, together they created the backbone of the city and provided the muscle pushing Worcester into the present.

Many people today wish to know more about the city, its history, and its early residents. A great deal of the information they seek can be found in one place – Worcester Historical Museum (WHM). Among the most regularly consulted books in WHM's library are Worcester City Directories ranging from 1870 to 1960. Within these volumes lies a wealth of information about vital statistics, government, businesses and common residents waiting to be mined by professor scholars, local historians, and genealogists. Where people lived, where they worked, the state of their economy, and ads for their businesses are all recorded in these aging books. Each city directory essentially offers a snapshot of the past and the people who made the city.

The utility of these directories to a variety of researchers is revealed by signs of heavy usage such as broken spines, chipped edges and loose pages. It soon becomes evident after brief examination of the physical state of the directories that preserving them is of utmost importance. However, they must at the same time be easily accessible to the public, because they are some of the most referenced items by genealogists, local history enthusiasts, and professional historians.



The best way to do this is through the quickly evolving field of digital preservation. Digitizing Worcester's history would be greatly beneficial to not only WHM and its collection, but to the average curious citizen, scholar, and city official.

## 1.1 Mission

The mission of this Interactive Qualifying Project is to assess, digitize, and make accessible WHM's collection of directories in a manner that presents the data in an effective and efficient interface for a diverse audience of museum patrons. This will involve:

1. Creating a protocol for handling and scanning the directories;
2. Translating the scanned Adobe files searchable through dirty optical character recognition;
3. Implementing a data storage and security plan; and
4. Investigating commercial and open source data management software platforms suitable for making the scanned directories available for easily accessible and searchable for museum patrons.

## 1.2 Values

The following core values and beliefs will guide progress toward the core mission:

1. Make broadly accessible WHM's city directory collection.
2. Create a digital collection that develops user understanding of the value of city directories, their utility and historical significance.
3. Provide a stable, scalable and sustainable platform for the delivery and long-term management of digital content.
4. Strive to deliver content of the highest quality in the ever-evolving best practices in digital history.

5. Present material using the appropriate metadata for integration into existing practices.
6. Take advantage of the assistance of available expertise from the various professionals working in information science, digital humanities and library sciences.

### 1.3 Problem Identification

The museum needs a system that bridges the gap between the directories and the patrons trying to access them. The patron could be a professional researcher, a student researching the population growth of Worcester during a specific time period, or just a curious homeowner who wants to know who lived on their property 150 years ago. This system should balance the various needs of patrons and the interests of the Worcester Historical Museum.

Every patron accessing the directories can be sorted into one of three categories. The first group deals with specific searches, where a person has a specific piece of information they seek. In this case a specific family, person, or other piece of data is the subject of a search to enrich a family history or community study. Someone who might want to trace a family tree or perform a genealogical search would fall under this category. The second is professional scholars who might aggregate data drawn from multiple directories to measure economic growth, industrial distribution, or racial and ethnic segregation. The third category includes the person who is visiting the museum and just so happens to find themselves going through the directory out of fascination. They may be more interested in broad searches or interesting trends. For example, someone may be interested by following a shift in the ethnic composition of the neighborhood they live in. These three different categories pose different needs that must be fulfilled.

These needs can be derived directly from the definition of each group. The ideal situation for the first group described would be an electronic, easily accessible and searchable, compilation of the directories. The second group requires a simple, easy to use, intuitive

interface that will encourage random interactions with the directories. However, currently these needs are not being fulfilled. The 1892 directory is currently the only one available. It is not easy to access, use, or search. Nor does it have any guidelines for use or encouragement of interaction. This directory was scanned and set up by a previous WPI Interactive Qualifying Project team in 2014. They were able to create a methodology for scanning a fragile directory. They were successful in collecting the scans and compiling them into one location, resulting in the Museum's one digital directory. However, the interface that the directories are presented in presents several limitations. Searching for keywords is difficult and tedious. Also, there is currently only one directory so there is no current methodology on how to searching multiple directories simultaneously. There is no integration into the exhibits so there is no encouragement of interaction from people passing by a computer terminal in the back of a gallery. These problems must be addressed if this project is to be successful.

Once additional directories are scanned, data must be presented in a manner that is not only easy to use and search through, but also promotes interaction. You must be able to search through multiple directories simultaneously and have the search be comprehensive and relevant. The digital directories must also be appealing to people in the gallery so they may feel encouraged to explore it. They should be easily accessible and available online as well.

## CHAPTER 2: ENVIRONMENTAL REVIEW

There are already plenty of examples of digitized directories and similar local history centers that utilize various programs and user interfaces to make the information accessible. Looking at what other cities have done with their directories, it becomes clear that there are plenty of viable options to choose from for presenting the data. This makes the selection of data management and search platforms extremely important along with the evaluating of the programs proprietary and open source abilities. When selecting the third party software program, the core needs of the different types of user's must be kept in mind and will drive the search for an ideal presentation of information.

### 2.1 Scholarly Literature

The challenge and potential benefits of digitizing city directories has been explored before. Numerous different individuals and institutions around the world have documented their experiences with bringing city directories online for easy public consumption. These experiences can be used to identify the best practices and potential hurdles.

Many different institutions around the world have evaluated and realized the benefits of digitizing city directories. The Library and Archives Canada (LAC) chronicles their current methods for preservation and distribution of both digital and physical documents. For example, if there are multiple copies of a physical document, then the copies are kept in two different locations as to cut down on the chance of the document being lost completely in an accident. These steps are taken because the LAC has recognized that city directories are a popular resource, and thus preservation is crucial. Digital preservation is also used by the LAC, due not only to its preservation capabilities but also its ability to make information widely available.

Consequently, the LAC pursued digitization of its city directories, and all 15 directories in the Canadian collection are searchable and available for public use (Maxwell, 2005).

The significance of digitizing historical documents was also examined by Professor Austin Kerr with regards to the research done by business historians. Professor Kerr, in an article written in 2000, recognized the potential of digital history. Digitized historical documents are shown to aid in research, “Perhaps most important about the digital revolution in historical data are the opportunities provided for answering new questions and for answering old questions more efficiently and effectively (Austin Kerr, 2000)”. Two examples of projects directed towards digitization of city directories are, Primary Source Media, a company interested in digitizing city directories and placing them online for public use, and also the American Memory Project by the Library of Congress. The goal of the American Memory Project was to digitize historical advertisements and make them totally searchable to improve information discovery. Similar projects with similar goals are multiplying, for there is still a need for, and an eager academic audience for, digitized historical information (Austin Kerr, 2000).

Another key example of city directory digitization is the experience of the McMillan Memorial Library in Wisconsin Rapids, Wisconsin. This library set out to digitize its city directories and observe the results. It first realized how valuable this information could be, stating, “By making readily available the raw material with which they are most conversant – the historical detail of their own communities – libraries can provide the building blocks for historians of all stripes and training to satisfy a wide range of personal and professional information needs (Litzer & Barnett, 2004)”. This still holds true today, and is applicable for all cases of digital history.

The McMillan Memorial Library used a combination of scanning and hand typing to digitize. Their first project was digitizing the book, *100 Years of Pictorial and Descriptive History of Wisconsin Rapids*. Some material of this book had already been transcribed using a typewriter, and this material was scanned in and used with OCR software. The remaining material was hand-typed using Microsoft Word, where this content was then converted to HTML format for web use. A digital scrapbook was created of the images by scanning them in the GIF and JPEG image formats. The library then continued its digitization project, moving on to more local artifacts, such as six Wisconsin Rapids city directories. These artifacts were digitized with funds from a grant the library received from the Consolidated Papers Foundation of Wisconsin Rapids. The grant money was used to hire a company that digitized materials as TIFF files and put them on CD-ROMs and on the web (Litzer & Barnett, 2004). This method fits similarly to what the previous IQP did, where they used a scanner and Adobe software to digitize a city directory and make it searchable (Fuat Becan, Harrington, Lundgren, & Orfan, 2014).

Following its digitization, McMillan Memorial Library tabulated the use of its digital history service. Web data was used for the analysis, however the data was not used raw. A very particular analysis was employed to attempt to weed out as many uses of the service that were not a true use, “Due to WebBots, spiders, in-house use, and other factors, statistics are generated that have no bearing upon actual research or other patron use (Litzer & Barnett, 2004)”. Techniques such as focusing on a specific period of time for analysis and concentrating only on the most used section of a multi-part document were used to compensate for inaccuracies and create useful data. The analysis showed that the digitized artifacts were being used by the public, thus justifying the process of getting these resources online (Litzer & Barnett, 2004).

In conclusion, the benefits of digitizing city directories have been examined and outlined by numerous individuals and organizations, both academic and community minded. The greatest profit of digitizing city directories comes from their aid in research. A repository of digitized city directories provides a large database of historical documents with easy search to the widespread and geographically dispersed public.

## 2.2 Digital Collections and Best Practices

On the Internet today there are many examples of digitized city directories. Many of these services also include documentation on their procedures and standards. Therefore, a lot can be learned from previous digitization efforts and can be applied to the research for the Worcester Historical Museum city directory digitization.

City directory digitization projects have been performed before and their features have advanced greatly over time. An example of an early system is that employed by the Tufts Digital Collections and Archives for the Boston City Directories. The scanning of the city directories used in this system was performed between 2002 and 2004 ("Boston Streets: Mapping Directory Data: FAQ," 2014). Here the directories are published as text, and are navigated by using an alphabetical list of all the names that appear in the directory. A name can be chosen, and then more information is shown. The system features a comprehensive search function, and provides a wealth of easy to find information to the inquiring historian. ("Boston Streets: Mapping Directory Data: City Directories," 2004).

As technology advanced, so did the capabilities of digitized city directories. With the Boston City Directories system, images were scanned, and then certain images were selected to be converted to text files("Boston Streets: Mapping Directory Data: FAQ," 2014). However, the Pasadena Digital History Collaboration system is powered by software called CONTENTdm

("Entering Subject Headings in CONTENTdm | Pasadena Digital History Collaboration," 2014).

This software provides many features, a prominent one being Optical Character Recognition (OCR). This allows the text from image files to automatically become searchable by the user of the database ("CONTENTdm Features, "). The Pasadena Digital History Collaboration also features a graphics rich city directory collection. The collection can be explored using the graphics of the covers of each city directory along with images of each of their pages. Like the Boston City Directories, the entire directory can be easily searched to locate necessary information ("Pasadena City Directories, ").

A table highlighting the qualities of several selected digitization project systems is shown below. The quantity field evaluates how many city directories are provided by the service. The User Interface (UI) field evaluates how the system is designed to promote ease of use and discovery. The search field looks at qualities such as a simultaneous search of all the directories and highlighting the results in the document. The speed field is based on how responsive the system is to user input and how quickly results are generated. Finally, the metadata field is based upon how much metadata is provided with the documents and how thorough it is. It is from observing examples like these that we can learn the best way to present the digitized city directories, in terms of both form and function.

	Pasadena	[1]Buffalo	Boston	[2]Pittsburgh	[3]Ancestry.com	[4]Butte County	[5]UNC
Quantity	✓	✓	✓	✓	✓	✓-	✓
UI	✓	✓+			✓		✓
Search	✓	✓	✓	✓	✓		✓



	Pasadena	[1]Buffalo	Boston	[2]Pittsburgh	[3]Ancestry.com	[4]Butte County	[5]UNC
Speed			✓	✓	✓	✓	✓
Metadata	✓	✓		✓-	NA		✓
Service	ContentDM	ContentDM	Tufts Fedora System	University of Pittsburgh Digital Research Library	Ancestry.com	Unknown	UNC Chapel Hill University Library

Table references: [1] "Buffalo & Erie County Public Library - Buffalo City Directories," [2] "Historic Pittsburgh City Directories," 2014, [3] "U.S. City Directories, 1821-1989," 2014, [4] "The Butte County Directories," 2014 [5] "North Carolina City Directories,"

Many of the systems that collect and make available these city directories have a series of practices and protocols that define their system. These systems can be used as inspiration to create guides for the digitization of the city directories for the Worcester Historical Museum. In an article written by Jessica and Gregory Colati (2011), there is an outline of basic features necessary for any digital repository, including:

Metadata creation – this information provides descriptions that allow for easy discovery and organization of data;

Usage controls – this ensures only those with proper permission have access to files;

Security – this ensures that information is safe both digitally and physically;

Disaster Recovery – in the event of lost data, this creates a plan to prevent the loss of data and a protocol for its recovery in the event of a disaster; and

Migration and Preservation Planning – this is making both the hardware and software used for the repository future-proof.

Following this list of criteria should provide the basis for a robust digital repository to house the city directories for the Worcester Historical Museum (Colati & Colati, 2011).

Another aspect of building a digital repository is ensuring the standards used for documenting the data are widely understood. There have been a series of standards developed over the years, and are outlined in another article by Jessica and Gregory Colati (2011). A digital repository needs to support compatibility with these standards to be able to be widely used. Metadata must be made compatible with many different systems, and this compatibility is defined by the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH). This set of standards outlines how metadata can be made to be compatible across all different access methods. In terms of how the software itself should be structured for use, the Reference Model for an Open Archival Information System (OAIS) provides a guide. This model outlines methods for creating a digital archive designed for long-term use, and the model is still in use today. The next stage in repository creation comes from the *Framework of Guidance for Building Good Digital Collections* (Framework). This framework provides four core principles, - Collections, Objects, Metadata, and Initiatives - that are designed to guide the creation of a digital archive, and adherence to these principles will generate a quality archive. The Framework is maintained by the National Information Standards Organization, (Branco Colati & Colati, 2011) an organization that also stores and maintains a large collection of standards (NISO Standards Projects, 2014) that can be referenced by all sorts of different organizations to attempt to bring unity to their formatting. This is best summed up the NISO homepage, “NISO is where content publishers, libraries, and software developers turn for information industry standards that allow them to work together ("Home - National Information Standards Organization," 2014)”.

When a large database of directories is created, there must be a way to search in an organized fashion. This is accomplished through the use of metadata, a system of parameters that are used to describe the contents of a document. Metadata is going to be necessary to organize

the city directories that are going to be scanned. A prominent metadata system is the Dublin Core. The Dublin Core is designed to be a flexible system for many different needs ("DCMI Metadata Basics," 2014), and is used in services such as Content Pro by Innovative Interfaces (Content Pro Fact Sheet). Other sources that can be used for help with metadata formatting are the Pasadena Digital History Collaboration Site and the Federal Agencies Digitization Guidelines Initiative (FADGI). The Pasadena Digital History Collaboration outlines the methods they use on their own site for storing and organizing city directories, such as their Rules, Policies, and Standards and their Subject Heading Guidelines ("Sitemap | Pasadena Digital History Collaboration,"). The FADGI is designed to "define common guidelines, methods, and practices to digitize historical content in a sustainable manner" ("About - Federal Agencies Digitization Guidelines Initiative," 2010). These sources can be used to establish a comprehensive system for organizing and classifying the city directories.

A system is going to be necessary to manage all of the city directories. The sheer quantity of directories makes their management difficult, thus requiring the use of a dedicated system. Two tools are CONTENTdm and Content Pro. Both services are designed to make it easy to input files and their respective metadata, along with making them easily searchable by the end-user. Since both products offer a similar set of services, either could potentially be used for storing the WHM's city directories. CONTENTdm was used for the Boise State University Western Writers Series collection, and provide the features similar to what is necessary for digitizing the city directories. The system offered a comprehensive search function, an easy to use interface, and metadata support ("Western Writers Series Digital Editions,"). Other content management systems, both commercial and open source, exist and can be explored for the purposes of managing the Worcester Historical Museum's city directory archive.

In terms of specific strengths and weaknesses, at this stage the content management systems are difficult to evaluate. CONTENTdm and Content Pro are advanced programs and consequently are best evaluated through use. A potential weakness of these programs is their usability, not for the end-user, but rather for the curator, specifically the Worcester Historical Museum. While both CONTENTdm and Content Pro brand themselves as easy to use for all users, this is something that would be best evaluated through use, and thus could be a potential downfall dependent upon user opinion. Also, neither system initially makes their price clear, and thus this could be a roadblock.

It can be seen from the existing examples of digitized city directories that there is an established system of best practices for their implementation. A lot can be learned just by observing how these systems function and what their guidelines for development were. These lessons can then be applied to the city directory digitization project of the Worcester Historical Museum in order to create the best result, both for the museum and the consumer.

### 2.3 Applications in the Digital Humanities

Digitized city directories, and digital history in general, has many different applications to historians of all levels, from amateur to professional. City directories are not only used for learning about a specific person or building, but they can also be used in projects. A project performed by the Emory University Libraries aims to take historical information from all sorts of different documents and map it geographically for historians to use, stating “Historical maps, atlases, and gazetteers are invaluable resources for scholars who examine phenomena with a space-time dynamic. Today’s geospatial technologies offer opportunities to link and transform spatial data from manuscript collections repositories and other collections within libraries” (Page, Durante, & Gue, 2013). One of the sources being used in the development of this project

is a collection of digitized city directories. The Atlanta City Directories were digitized using a Kirtas machine, and then OCR was used to extract the text that way the information could easily be used to contribute to the project (Page, Durante, & Gue, 2013).

A history course taught in 2014 by Lincoln Mullen at Brandeis University set out to map the history of religion in Boston. This was performed through the study of the various different congregations that existed throughout Boston over its history. A wide variety of sources were addressed over the course of the project, including “city directories, maps from the nineteenth century, newspapers, photographic collections...” (Mullen, 2014). Metadata played a key role in the project, for as students gathered more data a method of organizing it all quickly became necessary. While metadata was crucial, digitized historical documents were invaluable. The city directories used by the class had been digitized by the Google Books, the HathiTrust, and the Internet Archive. Since travel is difficult, this digitization is what made the class possible, “A class like this would not be possible apart from such digitization, because undergraduate students rarely travel even to nearby archives and libraries to use original materials” (Mullen, 2014).

## CHAPTER 3: METHODS

In creating a digital archive of Worcester City Directories, this project focused on creating standards and best practices in generating digital scans, saving them as workable PDFs; formatting them properly for OCR keyword search, and creating secure storage protocols.

### 3.1 Scanning

Although a feed scanner offers the promise of speed, this team ran into too many problems with jams to warrant its use. While not as efficient as a properly functioning feed scanner, the cradle scanner can be operated quickly and consistently with an effective procedure. Methods such as adjusting the pages, raising or lowering the supports under the directory, and using a blank page next to a cover page to help get a clear image of the cover are all necessary to use the cradle scanner. It is due to this risk and the consistent effectiveness of the cradle scanner that it is recommended to rely on the cradle scanner. When using the cradle scanner, it is recommended the following steps be followed in order to ensure an easy and reliable scanning process:

1. Determine type of scanning that needs to be done (Bed vs. Feed Scanning) Use bed scanner, much more effective
2. Schedule time with Archives to scan in the Archives
3. Place book gently in the scanner
  - a) Face up for 2 page scans at a time
4. Gently push down on the book with the glass cover
5. Press the green button or the “Scan” button on the computer
6. Make sure pages come out clearly and not corners are cut
  - a) If not, rescan until presentable

7. Flip Page
8. Repeat Steps 3-7 until you scan 20-30 pages
9. Save
  - a) Save original files with a name indicative of the contents of that file
  - b) Save onto thumb drive
10. Shutdown Program and log out of the computer
11. After leaving Archives, upload PDF files onto the WPI server

Through experience it was determined the best method to operate the cradle scanner is by employing three people. One person operates the computer that takes and saves the images of the pages, another turns each page as it is scanned, and the third operates the cover that presses the document down before it's scanned.

### 3.2 Storage

Once the city directories are scanned and saved, it is important to store them in multiple places to keep them protected from data loss. For this project city directories were stored on a WPI server available to students pursuing projects on campus. This is a relatively simple process were the computer with the city directories connects to the server, and can then simply copy the city directories from the computer's storage to the server storage. The one factor that is essential, is ensuring you have a consistent and reliable internet connection, without that it is almost impossible for the scans to be uploaded. Using an Ethernet cord is suggested because it allows for fast connection. The uploads vary from directory to directory, as a bigger directory has a larger file size, and thus takes longer to upload.

### 3.3 Image Manipulation

For after scanning, we save the pictures with 24 bit color, so that we have the ability to change that later on (Colati, 2014). It allows us to change them to black and white if that's what the Worcester History Museum wants. We also can use Photoshop to clean up the image and make it better quality. From the small sample size that we've seen, Photoshop might not be necessary, but we'll find out more once we have scanned more directories. The quality of the bed scanner is very clear and legible, so it won't require Photoshop while the feed scanner might have a different quality of scanning. Right now, we have an understanding of how to scan the directories, but we need more time to figure out a timeline of how many we can do and will figure out which directories will need to be scanned differently.

Various different digitization projects have had different results depending on what their topic was. As stated by the UNC digitization guides, we do minimal manual image manipulation when following the best digitization practices (Digitization Guidelines). The reason for not messing with these images is to prevent problems such as rotated or skewed images.

### 3.4 Optical Character Recognition

Once the scanning process is complete and the files have been saved, it is necessary to perform OCR. This process can be done in Adobe Reader or Adobe Acrobat XI. Acrobat was chosen because of its easy accessibility, as all WPI computers have access to the Acrobat software through an institution license. This made performing OCR as easy as taking the scans off the server running the OCR in only 3 steps. OCR on Acrobat is recommend because it is simple and it will be better to use the same program for all OCR.



## **OCR steps for adding to a scanned PDF document**

1. Open the PDF file.
2. In Acrobat, select View > Tools > Text Recognition. The Text Recognition panel in the Tools pane opens.
3. Click In This File. Designate the desired pages and click OK. Acrobat applies OCR to the scanned document.

The current state of the project is such that nine city directories are scanned and saved in multiple locations, and all have OCR performed on them.

## **List of Scanned Directories**

The following is a complete list of Worcester City Directories that have been digitized and their source.

1845- Digitizing Worcester's History (IQP)

1850- Digitizing Worcester's History (IQP)

1855- Digitizing Worcester's History (IQP)

1855- Google Books

1860- Digitizing Worcester's History (IQP)

1870- Digitizing Worcester's History (IQP)

1880- Digitizing Worcester's History (IQP)

1890- Digitizing Worcester's History (IQP)

1892- Worcester Historical Museum: An Interactive Redesign of the Fuller Gallery (IQP)

1920- Digitizing Worcester's History (IQP)

1925- Digitizing Worcester's History (IQP)

## **CHAPTER 4: DATA MANAGEMENT SYSTEMS**

Just one city directory is composed of multiple large, searchable, PDF files. Therefore a system is necessary to make these directories manageable for both the administrator and the end-user. The system needs to facilitate metadata, multiple files, and the ability to search across multiple city directories all at once, all while being easy to use for the WHM and their patrons. Multiple digital content management services were explored, with the five options being ContentDM, Omeka, Greenstone, ContentPro, and SimpleDL. These services all support metadata and simultaneous searching; however, each has different features. In terms of the important factor of cost, Omeka and Greenstone were free without hosting services, and ContentDM, ContentPro, and SimpleDL were paid. After sitting down with the WHM, it was determined that their preferences were ContentDM, Omeka, and SimpleDL, and that these services would be explored further and trialed. Omeka is a program that is growing in popularity, and was included for consideration due to the WHM's desire to at least investigate what could become a common program. ContentDM was considered because of its reputation, as it's used for many similar applications and is feature rich and well supported ("CONTENTdm Digital Media Management Software," 2015). SimpleDL was considered as a paid alternative to ContentDM, also offering the customer support a paid program includes ("SimpleDL. Digital Libraries Simplified," 2015). Greenstone was ruled out for being similar to Omeka, but without the valuable growing popularity, and ContentPro was ruled out due to difficulty contacting them to get purchase or trial details.

### **4.1 Trial One - ContentDM**

ContentDM is a widely used commercial service. ContentDM has all the desired features, including search, hosting services, support for multiple file types and metadata. ContentDM also

offers customer support for their users, making it easy for an institution to get assistance with any difficulties. ContentDM sites tend to be aesthetically pleasing and fully featured. Purchasing ContentDM is done through evaluating different price tiers and selecting the most appropriate. These tiers have different item limits, different project client count limits, and different OCR project client count limits, and it's important the institution determine which is appropriate for their needs. From there, the tier can be increased as the institution's collection grows. All these features command a hefty price however, restricting ContentDM to well-funded institutions ("CONTENTdm Digital Media Management Software," 2015).

An example of a ContentDM city directory setup is the Pasadena Digital History Collaboration City Directories collection. Users are presented with a digital carousel of screenshots of the city directory pages. Here they can choose a city directory they would like to explore, opening a new page that shows the PDF images of the pages in that directory. The user also has the option to switch to a text view, that shows just the text extracted from the PDF of the directory. Relevant metadata is shown directly underneath the content. Throughout the experience a search bar is available at the top of the window, allowing the user to search the entire collection at any time. The search will identify all the files that contain the search term, and then selecting the file will show all the instances of the search term in that file. The PDF view will not show the search term highlighted; however, switching to the text view will highlight where the term is in the document ("Pasadena City Directories,").

## 4.2 Trial Two - SimpleDL

Another commercial service is SimpleDL. SimpleDL, similar to ContentDM, offers hosting services, multiple file types, customer support, search, and metadata. Simple DL offers multiple options for plans, each with different storage capacities. In addition, SimpleDL offers a

customizable plan option, with a specific desired feature set for a custom price. SimpleDL is presented as another option to ContentDM, in the event the institution tries out the services and finds SimpleDL preferable ("SimpleDL. Digital Libraries Simplified," 2015).

SimpleDL is employed by Henderson Libraries of Henderson, Nevada to power their mypubliclibrary.com site, and provides a simple interface for the user. Upon entering their digital collection a page showing some of the items in their collection is shown, and along the left side bar different search parameters are available, such as collection, sub-collection, genre or subject. The user can search in the search bar, and the results are shown on screen, and the search parameters are still present to refine the search. When a result is selected, excerpts that contain the search term are shown in the left sidebar, and can be selected to jump to the excerpt's location in the document. The term is not highlighted on that page; however, the system brings the user directly to the page with the term ("mypubliclibrary.com Digital Collections,"). This is not consistent for all items, as some items are media based and the feature is not available. A video of an interview with Joe Hill that had a transcript did not show location of the search term in the transcript, however the transcript was a downloadable PDF, and therefore could be searched with a PDF viewer (Hill, 2014).

### 4.3 Trial Three - Omeka

The final recommended option is Omeka. Omeka is a free service that is supported through documentation and forums. These support options should be noted, as the lack of formal customer support may be a disadvantage, depending on the WHM's needs. Despite this, Omeka does offer the necessary features of hosting support, wide file type support, metadata support, and the ability to search. However, some of these features come through the use of plugins ("Omeka," 2015). Hosting services can come from Omeka themselves, through Omeka.net

("Omeka.net," 2015), or through other hosting services such as Hexahost ("Hexahost.com - Web hosting provider," 2015) and Reclaim Hosting ("Reclaim Hosting | Take Control of your Digital Identity," 2015).

The Chicopee Archives Online hosts city directories through Omeka. Here the user browses the city directories by title, and then can click on one to see all the relevant metadata and access the file. Opening the file presents it to the user in the browser's PDF viewer. Throughout the experience a search bar is provided on the side that allows the user to search all the content. Searching will show all the items that contain the search term, and then opening the item will allow the user to search the document for instances of the term using their browser's search function ("City Directories - Chicopee Archives Online," 2015). Omeka provides the option to have a fully featured digital collection for a budget conscious institution.

#### 4.4 Beta Test of ContentDM, SimpleDL, and Omeka.

In order to best choose a service for making city directories easily available and searchable, it is recommended that trial periods be used to accurately evaluate the options. For this evaluation, it was determined by the WHM that ContentDM, SimpleDL, and Omeka would be the focus. Every institution looking to digitize city directories will have different needs and different technical and financial capabilities. For the WHM, the use of hosting services will eliminate the challenge of server maintenance. Cost is also a concern for the WHM, and the different pricing schemes of all three services should be evaluated. These services provide the features necessary to complete this project and potentially a similar one, and offer a trial in some capacity, allowing the WHM to evaluate their options.

Setting up a trial is a little different for each service. ContentDM has a trial function the user can request and gain access to a hosted ContentDM site with all features for evaluation over

60 days ("CONTENTdm Digital Media Management Software," 2015). SimpleDL also has user-request access for a trial, granting 30-day access to 10GB of storage or 1000 items ("SimpleDL. Digital Libraries Simplified," 2015). Both of these trials provide all the features necessary for the WHM's application, with full metadata and search support. Omeka offers a free Omeka.net hosting plan that provides 500MB of storage ("Omeka.net," 2015), however it doesn't include the plugin necessary for PDF searching ("Omeka," 2015). The \$49 a year plan provides the plugin necessary along with 1GB of storage to facilitate the directories used for the trial. Once the trials are registered the services are available to use just like a full-time customer, making it easy to evaluate them.

In evaluating the services, it is recommended that the WHM focus on six variables: usability, search function, relation to other digital assets, costs, security, and technical expertise. Each service accomplishes all of these functions, it's just a matter of determining which does it best for the institution's needs.

#### 4.4.1 Usability

Starting with usability and search function, each service is usable, but to varying degrees. Usability and search function are explored together here, because the main focus for this project is search abilities, and therefore their usability is directly tied to how easily they can be searched. While usability is highly subjective, it is the opinion of this IQP team that ContentDM and SimpleDL are the most usable. They both offer custom user interfaces designed for making historical documents easy to explore. The user simply searches for a term, and they are presented with the results. The search points the user to the PDF of pages containing the search term, and then the user can search again to find all the instances of the term in that document.

Omeka, on the other hand, offers a decent user interface that is aesthetically pleasing however is not the most usable. The user must use two search methods, keyword and exact match, to ensure they are getting the desired results. The keyword method searches for terms based on relevancy. To use the example provided by Omeka, if a user searches for “ancient warfare”, they would get results first for “ancient warfare” and then for “ancient” and “warfare.” For exact match, the service searches for exact matches of the search term in the document ("Managing Search Settings 2.0," 2015). Sometimes in Omeka terms that are in the city directory do not show up with results when keyword search is used. This may be due to terms that appear frequently in documents being dropped from search results (dgrande77, patrickmj, & sbarnessckls, 2015). Therefore both search methods must be used to ensure that a comprehensive search is performed. Once the search term is found, Omeka points the user to the PDFs that contain pages with the search term. The user must then open the PDF in their browser’s PDF viewer and manually search for the desired term. These extra steps make the usability and search function of Omeka inferior to the offerings of ContentDM and SimpleDL.

#### 4.4.2 Administration

For the administrator, ContentDM is the most difficult to setup digital assets in due to its immense feature set. So many options are provided for the administrator that it can be difficult determining which ones are necessary. However the basics of uploading a city directory and making it searchable are still straightforward. SimpleDL offers fewer features than ContentDM, and, therefore, its administration is simpler. SimpleDL offers a very straightforward interface for uploading directories, and metadata for multiple directories can be easily edited at the same time on a spreadsheet. Omeka offers a user-friendly interface that makes it easy to upload city directories. Some actions require more steps than should be necessary, but overall the

administration interface is easy to understand. The technical expertise required to run these platforms is similar for all of them. Maintenance is taken care of by the hosting service, and therefore the only technical work that needs to be done by the WHM is the administration of the service, and as was discussed in the usability section, all three services are relatively easy to administer. The usability of the administration of these interfaces is important to evaluate since these interfaces will be used frequently to operate the services. ("CONTENTdm Digital Media Management Software," 2015), ("SimpleDL. Digital Libraries Simplified," 2015), ("Omeka," 2015).

#### 4.4.3 Cost

The remaining variables are more difficult to evaluate and are more dependent upon the needs of the WHM. Each service offers support for many different file types that could be useful for the digital assets the WHM or another institution are looking to make available digitally. The costs of each service vary greatly. A level 1 ContentDM license that includes hosting services and a cap of 10,000 items or 40 Gigabytes (GB) costs \$11,813 for the first year. After the first year the WHM would only pay maintenance fees of \$3,372 per year (Hojnacki, 2015). For SimpleDL, the recommended starting package is the Share and Protect 2k. This allows 2000 files or 25GB, and costs \$3500 a year or \$292 a month. Finally, for Omeka, many different hosting services can be used. For Omeka.net hosting, the Platinum plan would be required for the 25GB of storage offered, costing \$999 a year ("Omeka.net," 2015).

#### 4.4.4 Security

Security for these services is difficult to comment on, as the details are not well advertised. Both ContentDM and SimpleDL advertise backup and monitoring of digital assets uploaded, however Omeka does not seem to comment on security.



## CHAPTER 5: FINDINGS AND RECOMMENDATIONS

### 5.1 Findings

**Finding 1: When scanning, the best option is to use the cradle scanner. While the feed scanner can be effective, the cradle scanner is the most reliable option.**

The feed scanner can have difficulty automatically loading the pages, and sometimes the pages get caught inside the mechanism. This results in the scanner jamming, and the process must be stopped in order to clear the jam. For certain directories, this is not much of an issue. Simply altering how many pages are loaded in the feed bay can minimize the jamming, and these directories can be scanned easily. However, for other directories the feed scanner will not work, as any amount of pages will result in the first one in the pile being jammed, and the machine is rendered ineffective. Furthermore, from our observation we believe the ability to scan pages in the feed scanner is related to the size and brittleness of the page, as some pages would work while others wouldn't. Therefore, the inconsistency associated with the feed scanner makes it difficult to use.

Once a book's binding is broken and the pages separated, it becomes difficult to use the cradle scanner. Without a binding, the book is harder to keep steady and organized while scanning. Therefore, it becomes a risk to try the feed scanner, as breaking the binding will make cradle scanning difficult should the feed scanner not work. In addition, while not as efficient as a properly functioning feed scanner, the cradle scanner can be operated quickly and consistently with an effective procedure. Methods such as adjusting the pages, raising or lowering the supports under the directory, and using a blank page next to a cover page to help get a clear image of the cover are all necessary to use the cradle scanner. It is due to this risk and the

consistent effectiveness of the cradle scanner that it is recommended to rely on the cradle scanner.

**Finding 2: During scanning, it was discovered that the PDF format is the best to rely on for scanning the city directories.**

It was found that this format does not use much space on computer storage or server storage, and can easily have OCR performed on it for searching later on. This is opposed to the TIFF format, where images take up a considerable amount of storage. In addition, PDF files count as only one item against the item cap for a ContentDM level. Regardless of the number of pages in a PDF file, the file counts as just one item for ContentDM. Therefore, if an institution is looking to use ContentDM, this is important to consider, as each level has different item caps, and there is the potential to save money by putting multiple pages of a large city directory in just one page ("CONTENTdm Digital Media Management Software," 2015). Regardless of the file choice, it is important to store the files in multiple locations to prevent from malfunction or other unexpected events compromising data.

**Finding 3: Dirty Optical Character Recognition provides sufficient accuracy for the purposes of this project.**

OCR is a very simple process that allows for the PDFs to become searchable using the control F function. There are two types of OCR, dirty and clean. The type that Acrobat does is dirty because it does a whole document search and picks out the easy words. Clean OCR is when a professional company goes through every page and makes sure each word is correctly spotted and added to the inventory. For our scans, we did dirty OCR because it makes more sense for us to do that over a 1,200 page book. Also, for the trial that we presented to the Worcester History

Museum, that worked well. In the future, they may decide to get clean OCR done, but for now there's no need to go into that much depth.

## 5.2 Recommendations

**Recommendation 1: When using the cradle scanner, it is recommended the following steps be followed in order to ensure an easy and reliable scanning process.**

### Bed Scanning Process

1. Determine type of scanning that needs to be done (Bed vs. Feed Scanning) Use bed scanner, much more effective
2. Schedule time with Archives to scan in the Archives
3. Place book gently in the scanner
  - a) Face up for 2 page scans at a time
4. Gently push down on the book with the glass cover
5. Press the green button or the "Scan" button on the computer
6. Make sure pages come out clearly and not corners are cut
  - a) If not, rescan until presentable
7. Flip Page
8. Repeat Steps 3-7 until you scan 20-30 pages
9. Save
  - a) Save original files with a name indicative of the contents of that file
  - b) Save onto thumb drive
10. Shutdown Program and log out of the computer
11. After leaving Archives, upload PDF files onto the WPI server

**Recommendation 2: Once the city directories are scanned and saved, it is important to store them in multiple places to keep them protected from data loss.**

For this project city directories were stored on a WPI server available to students pursuing projects on campus. This is a relatively simple process where the computer with the city directories connects to the server, and can then simply copy the city directories from the computer's storage to the server storage. The one factor that is essential, is ensuring you have a consistent and reliable internet connection, without that it is almost impossible for the scans to be uploaded. Using an Ethernet cord is suggested because it allows for fast connection. The uploads vary from directory to directory, as a bigger directory has a larger file size, and thus takes longer to upload.

**Recommendation 3: Once the scanning process is complete and the files have been saved, it is necessary to perform OCR.**

This process can be done in Adobe Reader or Adobe Acrobat XI. Acrobat was chosen because of its easy accessibility, as all WPI computers have access to the Acrobat software through an institution license. This made performing OCR as easy as taking the scans off the server running the OCR in only 3 steps. OCR on Acrobat is recommended because it is simple and it will be better to use the same program for all OCR.

OCR steps for adding to a scanned PDF document

1. Open the PDF file.
2. In Acrobat, select View > Tools > Text Recognition. The Text Recognition panel in the Tools pane opens.
3. Click In This File. Designate the desired pages and click OK. Acrobat applies OCR to the scanned document.

**Recommendation 4: From this point it is recommended that the WHM continue exploring the different software options presented.**

By doing so they will be able to choose the one that best fits their needs, now and for the future expansion of their digital collection. The current state of the project is such that nine city directories are scanned and saved in multiple locations, and all have OCR performed on them. A trial run of ContentDM, SimpleDL, and Omeka has been set up for the WHM to explore.

**Recommendation 5: Once this selection is made, the scanning process should continue in order to maximize the digital collection as quickly as possible.**

For this process, a directory for every ten years of history in the WHM museum should be scanned. Once the entire time period covered by the city directories has been scanned, the scanning team should then go through the collection again, now scanning directories with five years between their respective dates. This procedure should continue, with the gap between city directories closing for each scanning of the collection, until all the directories have been scanned. By using this procedure, it ensures that there is always a wide range of history covered by the digitized collection. If, for example, a team sets out to scan ten directories, and they scan directories from 1850 – 1860, this provides a small sample size of history to the user. However, by scanning ten directories from 1850 to 1950, it provides a wide range of dates to be explored. This procedure will result in the WHM having a rich collection of digitized city directories and software that is friendly to both the administrator and the user to explore.

## SOURCES

- About - Federal Agencies Digitization Guidelines Initiative. (2010). Retrieved 2014, 11/25/2014, from <http://www.digitizationguidelines.gov/about/>
- Austin Kerr, K. (2000). Revisiting business historians and the Internet. *Enterprise and Society*, 1(1), 179-181. doi: 10.1093/es/1.1.179
- Boston Streets: Mapping Directory Data: City Directories. (2004). Retrieved 11/24/14, 2014, from <http://dca.lib.tufts.edu/features/bostonstreets/people/directories.html>
- Boston Streets: Mapping Directory Data: FAQ. (2014). Retrieved 12/16/14, 2014, from <http://dca.lib.tufts.edu/features/bostonstreets/about/help/faq.html><http://dca.lib.tufts.edu/features/bostonstreets/people/directories.html>
- Branco Colati, J., & Colati, G. C. (2011). Road Tripping Down the Digital Preservation Highway, Part II: Road Signs, Billboards, and Rest Stops Along the Way. *Journal of Electronic Resources Librarianship*, 23(2), 165-173.
- Buffalo & Erie County Public Library - Buffalo City Directories. Retrieved 12/7/14, 2014, from <http://nyheritage.mnyn.net/cdm/search/collection/VHB011>
- City Directories - Chicopee Archives Online. (2015). Retrieved 5/2, 2015, from <http://www.chicopeepubliclibrary.org/archives/city-directories>
- Colati, J. (2014). Scanning B. Connolly.
- Colati, J. B., & Colati, G. C. (2011). Road Tripping Down the Digital Preservation Highway, Part I: Hitting the Road. *Journal of Electronic Resources Librarianship*, 23(1), 60-67. <http://www.oclc.org/en-US/contentdm/features.html>
- CONTENTdm Digital Media Management Software. (2015). Retrieved 4/12, 2015, from <http://www.contentdm.org/>
- DCMI Metadata Basics. (2014). Retrieved 11/24/14, 2014, from <http://dublincore.org/metadata-basics/>
- dgrande77, patrickmj, & sbarnessckls. (2015). Omeka for City Directories “ Omeka Forums. Retrieved 4/26, 2015, from <http://omeka.org/forums/topic/omeka-for-city-directories-post-106543>
- Entering Subject Headings in CONTENTdm | Pasadena Digital History Collaboration. (2014). Retrieved 12/16/14, 2014, from <http://pasadenadigitalhistory.com/partners/digitization/entering-subject-headings-in-contentdm/http://dublincore.org/metadata-basics/>
- Fuat Becan, A., Harrington, M., Lundgren, K., & Orfan, K. (2014). Worcester Historical Museum: An Interactive Redesign of the Fuller Gallery: Worcester Polytechnic Institute.
- Hexahost.com - Web hosting provider. (2015). Retrieved 4/12, 2015, from <http://hexahost.com/>
- Hill, J. (2014, 4/17/2014) *Joe Hill (2014-04-17) video oral history and transcript - Henderson Memories/Interviewer: A. M. Hamilton-Brehm*. Henderson Libraries.
- Historic Pittsburgh City Directories. (2014). Retrieved 12/8/14, 2014, from <http://digital.library.pitt.edu/cgi-bin/t/text/text-idx?c=pitttextdir>

- Hojnacki, S. (2015). OCLC CONTENTdm FY15 Quote - Worcester History Museum Library (UKO).
- Home - National Information Standards Organization. (2014). Retrieved 12/8/14, 2014, from <http://www.niso.org/home/>
- Litzer, D., & Barnett, A. (2004). Local history in e-books and on the Web: One library's experience as example and model. *Reference & User Services Quarterly*, 248-257.
- Managing Search Settings 2.0. (2015). Retrieved 4/26, 2015, from [https://omeka.org/codex/Managing\\_Search\\_Settings\\_2.0](https://omeka.org/codex/Managing_Search_Settings_2.0)
- mypubliclibrary.com Digital Collections. Retrieved 5/2, 2015, from <http://digitalcollections.mypubliclibrary.com/?q=&fieldLimit=&fq%5B%5D=q&fq%5B%5D=c2>
- Maxwell, D. (2005). Preservation of city directories at library and archives Canada. *The Serials Librarian*, 49(1-2), 205-210.
- Mullen, L. (2014). Using Metadata and Maps to Teach the History of Religion.
- NISO Standards Projects. (2014). Retrieved 12/16/14, 2014
- North Carolina City Directories. Retrieved 12/9/2014, 2014, from <http://www.digitalnc.org/collections/city-directories/>
- Omeka. (2015). Retrieved 4/12, 2015, from <http://omeka.org/>
- Omeka.net. (2015). Retrieved 4/12, 2015, from <https://http://www.omeka.net/>
- Pasadena City Directories. from <http://collection.pasadenadigitalhistory.com/cdm/landingpage/collection/p16237coll4>
- Page, M. C., Durante, K., & Gue, R. (2013). Modeling the History of the City. <http://dx.doi.org/10.1080/15420353.2013.767763>. doi: Journal of Map & Geography Libraries, Vol. 9, No. 1-2, January-August 2013, pp. 128-139
- <http://www.digitalnc.org/collections/city-directories/>
- Reclaim Hosting | Take Control of your Digital Identity. (2015). Retrieved 4/12, 2015, from <https://reclaimhosting.com/>
- SimpleDL. Digital Libraries Simplified. (2015). Retrieved 4/12, 2015, from <http://www.simpdled.com/index.html>
- Sitemap | Pasadena Digital History Collaboration. Retrieved 11/25/2014, 2014, from <http://pasadenadigitalhistory.com/sitemap/>
- The Butte County Directories. (2014). Retrieved 12/7/14, 2014, from <http://www.csuchico.edu/lspr/bcd/>
- U.S. City Directories, 1821-1989. (2014). Retrieved 12/7/14, 2014, from <http://search.ancestry.com/search/db.aspx?dbid=2469>
- Western Writers Series Digital Editions. Retrieved 11/25/14, 2014, from <http://digital.boisestate.edu/cdm/landingpage/collection/western>