

STAT210: Parsons Paper Company Payroll Register Curation

Kevin Jin and Nahia Pino

2023-03-02

Table of contents

1	Introduction	2
2	Processes to Curate Ledger	4
2.1	Scanning Process	4
2.2	Conversion Process	5
2.3	Process to convert .CR2 files to .tiff	7
2.4	Quality Assurance: Checking for Missing Pages	10
3	Viewing the Register	10
4	Next Steps: Adding to the Digital Commonwealth	11
5	Acknowledgements	12

1 Introduction

The Parsons Paper Copy Register consists of employee payroll and accounting records from January 1861 to April 1869 from Parsons Paper Company in Holyoke, Massachusetts. Founded in 1853 by Joseph Parsons, Parsons Paper Company was the oldest and largest manufacturer of cotton-based high quality writing papers in Holyoke until its liquidation in 2005.

The original pages of the register, which is now property of the Holyoke Public Library History Room, were scanned on January 10, 2023 as part of this class (STAT210: Mining the History of Holyoke) and screenshots are included in the analysis below. Pages contain information such as employee name and signature, role, days worked, daily pay, and total wages earned, and are organized by separate pay periods.

XX insert information about Shared Google Drive and Shiny app here XX

The original scanned files can be found on our Shared Google Drive here: <https://drive.google.com/drive/folders/18Aw57Hhga52E3skMttF7sxDfL6KWO-7q?usp=sharing>

For a current list of pages with issues, please look at the public Github repository here: https://github.com/STAT210-S23/Parsons_Paper_Register

Lastly, we created an interactive web app that can help display the converted scans. This interactive web applet can be found at <https://r.amherst.edu/apps/nhorton/Parsons-Paper>

It is important to note that the pages from the register had some interesting peculiarities—as it is common due to how old it is, the register is missing some pages, and it has cutouts on some pages where we believe may have contained revenue stamps before (as seen on Figure 1).

This document describes the process of scanning, transforming, and curating the images from the register and details how the register might be useful for historians and residents. Specifically, the links above can be used by historians and residents to more easily view this aspect of Holyoke's history:

- Original Scans: <https://drive.google.com/drive/folders/18Aw57Hhga52E3skMttF7sxDfL6KWO-7q?usp=sharing>
- GitHub: https://github.com/STAT210-S23/Parsons_Paper_Register
- Interactive Web Viewer of Ledger Pages: <https://r.amherst.edu/apps/nhorton/Parsons-Paper>

Received of the PARSONS PAPER COMPANY, by their Agent, the sum professed to our Signatures,
being in full of all demands up to October 1868

NAMES	No. POUNDS	No. REAMS	No. DAYS	PRICE	TOTAL AMOUNT	RENT, OR BOARD	BALANCE DUE	DATE	SIGNATURES
William Kelley			26	15	78.00	6.50	71.50		William Kelley
Michael Ryan			26	146	55.36	5.20	50.16		Michael Ryan
Ellen Conroy			26	77	30.33		30.33		Ellen Conroy
Kate Sheridan			26	"	30.33	7.00	23.33		Kate Sheridan
Ann Casey			26	"	30.33		30.33		Ann Casey
J. Mc Carthy			26	"	30.33		30.33		J. Mc Carthy
Mary Davenport			26	69	29.25		29.25		Mary Davenport
Ellen Keefe			26	69	29.25		29.25		Ellen Keefe
Sarah McQuibban			26	"	28.81		28.81		Sarah McQuibban
Mary Kelly			26	"	18.59		18.59		Mary Kelly
Mary Keenan			26	"	22.88		22.88		Mary Keenan
Julia Moriarty			26	"	15.28		15.28		Julia Moriarty
Louisa Quinlan			26	"	15.28	2.00	13.28		Louisa Quinlan
Mary Golding			26	"	22.28		22.28		Mary Golding
Kate O'Hare			26	"	26.33		26.33		Kate O'Hare
Ann Welch			26	"	20.16		20.16		Ann Welch
Bridget O'Connell			26	"	18.30		18.30		Bridget O'Connell
M. McQuibban			26	"	29.51		29.51		M. McQuibban
Alice Fennessy			26	"	25.05		25.05		Alice Fennessy
M. Rohan			26	"	22.07		22.07		M. Rohan
Bridget Fitzpatrick			26	"	22.87		22.87		Bridget Fitzpatrick
Ellen Shinn			26	"	28.27		28.27		Ellen Shinn
Bridget Kennedy			26	"	26.57		26.57		Bridget Kennedy
Mary Mullins			26	"	16.26		16.26		Mary Mullins
Mary Harkin			26	"	20.57		20.57		Mary Harkin
Mary Kelly			26	"	16.18		16.18		Mary Kelly
Ann Ryan			26	"	12.62		12.62		Ann Ryan
Mary Collins			26	7	16.32		16.32		Mary Collins
Ellen Maguire			26	7	21.93		21.93		Ellen Maguire
Ellen Sheehan			26	5	20.17		20.17		Ellen Sheehan
Mary Sheehan			26	5	29.14		29.14		Mary Sheehan
Hannah Harris			26	5	25.16		25.16		Hannah Harris
Kate Welch			26	5	15.07		15.07		Kate Welch
Kate Shea			26	5	18.05		18.05		Kate Shea
Ellen Burke			26	4	26.53		26.53		Ellen Burke
Kate Dunne			26	3	35.10		35.10		Kate Dunne
Mary Kelly			26	2	20.83		20.83		Mary Kelly
Bridget Healy			26	2	21.66		21.66		Bridget Healy
Hannah Sullivan			26	2	2.30		2.30		Hannah Sullivan
Eliza Lowe			26	2	17.76		17.76		Eliza Lowe
Margt. Bly			26	2	12.02		12.02		Margt. Bly
Mary Don			26	2	29.56		29.56		Mary Don
Kate McCarthy			26	2	16.61		16.61		Kate McCarthy
Rosey McCarthy			26	2	25.23		25.23		Rosey McCarthy
Ellen Shanley			26	2	22.63		22.63		Ellen Shanley
Ann K			26	1	32.12		32.12		Ann K
Kate			26	1	22.20		22.20		Kate
Mary			26	1	24.22		24.22		Mary
Julia			26	1	31.45		31.45		Julia
Bridget			26	1	25.96		25.96		Bridget
Hannah			26	1	32.49		32.49		Hannah
Julia			26	1	22.22		22.22		Julia
			26	1	15.40		15.40		



Figure 1: p. 286 of register after color correcting. On the left is a large rectangle cut off, a little bigger than the width of the stamps, suggesting they were cut off.

2 Processes to Curate Ledger

2.1 Scanning Process

Images of the register were scanned in January, 2023 using a Canon 5d Mark II with 24-105mm Canon zoom lens using the EOS Utility software on an Apple laptop.

Preview information:

- Color Model: RGB
- Depth: 16
- Orientation: 1 (Normal)
- Pixel Height: 3,744
- Pixel Width: 5,616
- Profile Name: Display P3

Canon information:

- Camera Temperature: 38
- Continuous Drive: Single Frame
- Focus Distance Lower: 0.79
- Focus Distance Upper: 0.84
- Lens Model: EF24-105mm f/4L IS USM
- Max Aperture Value: 4
- Min Aperture Value: 22.627
- Unique Model ID: 2,147,484,184
- White Balance Index: Auto

Exif:

- Aperture Value: 6.625
- Color Space: Uncalibrated
- Components Configuration: 1, 2, 3, 0
- Custom Rendered: Normal process
- Date Time Digitized: Jan 9, 2023 at 11:10:53 PM
- Date Time Original: Jan 9, 2023 at 11:10:53 PM
- Exif Version: 2.2.1
- Exposure Bias Value: 0
- Exposure Mode: Manual exposure
- Exposure Program: Manual
- Exposure Time: 1/4
- Flash: Off, did not fire
- FlashPix Version: 1.0
- FNumber: 10

- Focal Length: 50
- Focal Plane Resolution Unit: inches
- Focal Plane X Resolution: 3,849.212
- Focal Plane Y Resolution: 3,908.142
- ISO Speed Ratings: 200
- Max Aperture Value: 4
- Metering Mode: Pattern
- Pixel X Dimension: 5,616
- Pixel Y Dimension: 3,744
- Scene Capture Type: Standard
- Shutter Speed Value: 1/4
- Sub-second Time: 88
- Sub-second Time Digitized: 88
- Sub-second Time Original: 88
- White Balance: Auto white balance
- AFInfo: 0.379, 0.225, 0.1, 0.199, s
- Firmware: Firmware Version 2.0.7
- Flash Compensation: 0
- Focus Mode: AI Servo AF
- Image Stabilization: Panning
- Lens ID: 237
- Lens Info: 24, 105, 0, 0
- Lens Model: Canon EF 24-105mm f/4L IS

Picture style:

- Color Tone: 0, 0, 0
- Monochrome: 0, 0, 0
- Pict Style Color Space: Adobe RGB, 2, 2
- Picture Style: 4,095, 4,095, 4,095
- Sharpness Frequency: 0, 0, 0

Some additional images (to replace missing or blurry images) were scanned with an Apple iPhone 12 in February, 2023.

2.2 Conversion Process

Figure 2 displays a sample page from the register. These images were from raw camera files (.CR2) to .tiff files that we kept in the `tiff_original` directory in our Google Drive.

As some some pages had multiple scans (saved as `XXXa.CR2`, `XXXb.CR2`, etc.), we used bash commands in the terminal to identify any repeated images. For each page with duplicate scans, we inspected each version and selected the one with the best quality (based on quality

Received of the PARSONS PAPER COMPANY, by their Agent, the sum specified to our Signatures,
being in full of all demands up to February 1st 1867.

NAMES	No. POUNDS.	No. REAMS.	No. DAYS.	PRICE.	TOTAL AMOUNT.	RENT, OR BOARD.	BALANCE DUE.	DATE.	SIGNATURES.
William Johnson			24	166	67 37		67 37		Wm Johnson
John A. Egan			21	290	34 95	11 67	30 28		John A. Egan
William Connor			25	146	68 75	7 25	61 50		William Connor
Martin Kennedy			23	90	37 38		37 38		Martin Kennedy
James Casey			23	122	51 00		51		James Casey
Chas. Gallivan			25	0	37 50		37 50		Chas. Gallivan
John Vaughan			25	97	50 00		50		John Vaughan
Pat. Doyle			27	97	40 50		40 50		Pat. Doyle
Sam'l J. Wright			27	92	55 04	7 25	47 75		Sam'l J. Wright
Phillip Gidday			25	122	57 00		57		Phillip Gidday
John Conchen			25	97	56 00	6 25	49 75		John Conchen
Frank Russell			27	116	48 12	9 00	39 12		Frank Russell
Chauncy Johnson			28	16	35 63		35 63		Chauncy Johnson
William Connor			27	212	94 50		94 50		William Connor
John Flynn			27	192	67 50		67 50		John Flynn
E. W. Wellington			27	146	57 75	6 25	51 50		E. W. Wellington
Martin Moran			27	90	44 70		44 70		Martin Moran
Wm. M. Key			30	16	48 31	9 33	38 98		Wm. M. Key
Chas. Shing			27	132	58 50		58 50		Chas. Shing
J. H. Allen			27	97	40 50	5 00	35 50		J. H. Allen
Nick Ryan			31	14	56 53		56 53		Nick Ryan
Pat. Ryan			31	97	46 50		46 50		Pat. Ryan
Nick McKam			29	11	53 17		53 17		Nick McKam
Edward Wolf			25	16	40 63		40 63		Edward Wolf
Pat. Mayday			25	16	44 62	4 67	39 95		Pat. Mayday
Pat. O'Hare			27	90	48 58	5 70	38 58		Pat. O'Hare
Joe Beaudet			6	10	9 75	5 00	4 75		Joe Beaudet
Joe Beaudet Jr			25	90	45 50		45 50		Joe Beaudet Jr
John Naughton			27	116	47 25		47 25		John Naughton
Eugene Shanley			26	92	39 00		39		Eugene Shanley
William H. Davis			26	92	39 75		39 75		William H. Davis
John Shea			18	97	27 75		27 75		John Shea
John Parson			2	97	75		75		John Parson
Mary Inley			11	40	12 37		12 37		Mary Inley
Kate Kilham			27	49	30 37		30 37		Kate Kilham
Kate Moran			34	0	27		27		Kate Moran
Kate Braden			23	0	25 87		25 87		Kate Braden
Phillip Gidday					7 12		7 12		Phillip Gidday
John Flynn					1 00		1 00		John Flynn
Parsons & Co					1730 07	69 67	1660 40		
Austin Ely			25	157	76 50		76 50		Austin Ely
Walter Buckley			27	181	61 50		61 50		Walter Buckley
J. M. Alden			25	157	62 50		62 50		J. M. Alden
Martin Kennel			33	136	51 75	14 00	47 75		Martin Kennel
Pat. Casey			32	290	52 82		52 82		Pat. Casey
Austin Ely			27	19	81 00		81 00		Austin Ely
Walter Buckley			27	18	81 00		81 00		Walter Buckley
J. M. Alden			27	19	67 50		67 50		J. M. Alden
Martin Kennel			27	186	60 75	4 00	56 75		Martin Kennel
Pat. Casey			33	99	53 62		53 62		Pat. Casey
					342 87	4 00	338 87		

Figure 2: Ex. p. 253 of register - part of the February 1st, 1867 payroll

of photo, position of page, clarity of the words, and general preference). The rejected scans were moved into a separate folder labeled `raw_deleted`, in the Google Drive. If the scan remaining in `raw` was the second version, it was renamed to remove the additional letter.

2.3 Process to convert .CR2 files to .tiff

Within the Google Drive for Desktop app, using command selection every photo in the folder was selected and copy-pasted into the same `tiff_original` folder. These copies are automatically labeled as `Copy of tiff_XXX.tiff`. A new folder was created called `processing_tiff` and the copies were moved into this folder in batches of 10 using the command select function.

The process moved from the online Google Drive platform to the desktop version (which was downloaded). Within the desktop interface, 20 images were downloaded locally at a time. Starting from the image with the largest number, the image was opened with MacOS's `preview` application. In the application, first the image was cropped to remove excess space on the image. The size of the cropping was set based on making sure every part of the page itself was visible, see Figure 3 for an example. Dead space was kept in if it preserved parts of the page, for instance more bent pages would retain some dead space as the cropping was done in rectangle shapes, see Figure 4 for an example.

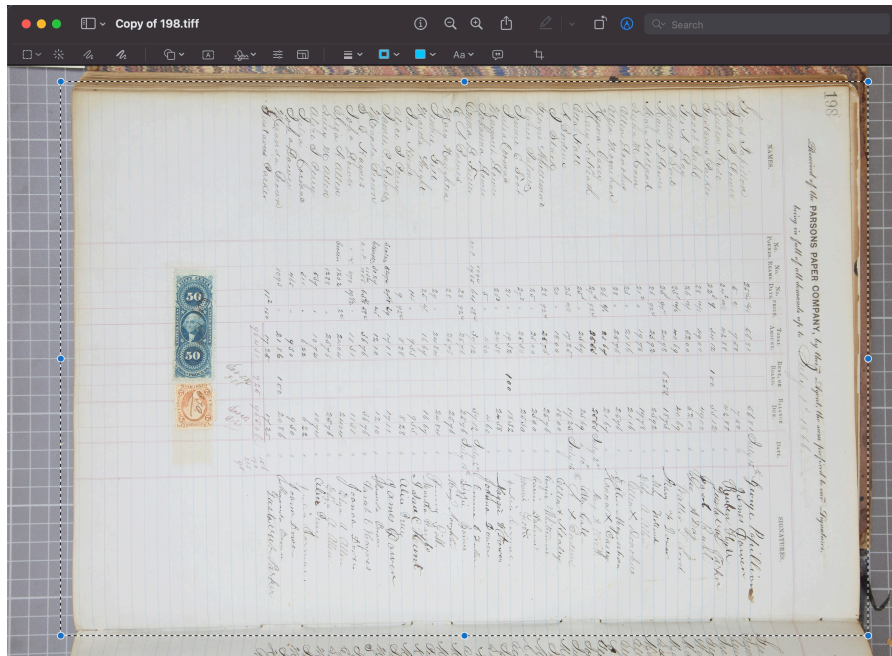


Figure 3: Ex. cropping p. 198 of register in preview

Received of the PARSONS PAPER COMPANY, by their Agent, the sum paid to our Signatures,
 being in full of all demands up to March 1st 1869

NAMES, JOHN	No. POUNDS.	No. REAMS.	No. DAYS.	PRICE.	TOTAL AMOUNT.	RENT, OR BOARD.	BALANCE DUE.	DATE.	SIGNATURES.
Geo Poppleton			24	11	72.00		72.00		George Poppleton
Walter J. Sords			16	13	41.25	7.25	54.00		Walter J. Sords
Wm Grant			23	12	57.19		57.19		Wm Grant
Geo A. Day			23	15	58.75	7.25	51.50		Geo A. Day
J. James			24	15	60.00		60.00		J. James
Chas H. Robinson			24	12	48.00		48.00		Chas H. Robinson
Ch. Morgan			24	7	28.58		28.58		Ch. Morgan
James Walker			24	10	42.00		42.00		James Walker
Frank Smith			24	9	36.00		36.00		Frank Smith
Anna Coxey			23	6	23.75		23.75		Anna Coxey
George Whitmore			11		11.00	6.00	5.00		George Whitmore
Mary Dowd			10		10.50	6.25	4.25		Mary Dowd
Thomas Baily			24		24.00		24.00		Thomas Baily
John Water			24		24	12	12		John Water
S. Buckland			24		24	12.00	12.00		S. Buckland
Emma Coulter			24		24		24.00		Emma Coulter
E. Dawson			23		23.50	12.00	11.50		E. Dawson
John Dart			22		22	12.00	10		John Dart
Mary Holland			23		23.00		23.00		Mary Holland
Maria Dart			23		23.80	12	11.80		Maria Dart
Ellen Kempton			23		23.80	12.00	11		Ellen Kempton
John Colby			22		22		22		John Colby
John Gunt			24		24		24		John Gunt
Israh Slater			24		24	12.00	12		Israh Slater
Anna Hill			24		24		24		Anna Hill
Anna Jones			24		24		24		Anna Jones
Felia Stockman			24	7	28		28		Felia Stockman
Maria Ross			52	6	6.60		6.60		Maria Ross
Sam Allen			39	48	53		28.33		Sam Allen
Wm. J. Brown			22	21	46		26.11		Wm. J. Brown
J. S. Mackel			20	6	29		16.24		J. S. Mackel
E. A. Grant			25	23	59		5.99		E. A. Grant
Maria Binnig			25	19	11	40	9.00	2.40	Maria Binnig
E. A. Allen			39	0	2		23.32		E. A. Allen
Ellen M. Gooden			25	13	65		13.00		Ellen M. Gooden
Ellen Miles			29	6	205		24.10	13.00	Ellen Miles
E. Phillips			26	3	231		26.20		E. Phillips
S. Conner			69	8	618		12.00	36	S. Conner
Martha Ross			24	1	916		18.32	12.50	Martha Ross
Louise Hill			11	14			23.28		Louise Hill
Mrs. Ellen Morgan			1	15	5		3.12	3.12	Mrs. Ellen Morgan
Mrs. Ellen N. Giffen				3	6		7.32	7.32	Mrs. Ellen N. Giffen
Chas. W. Giffen				12	0		0.00	7.00	Chas. W. Giffen
Mary Ross			8	6	3		17.26	17.26	Mary Ross
Frank H. Giffen			8	12	6		6.25	6.25	Frank H. Giffen
Geo. Poppleton			13	14	41		41.25	41.25	Geo. Poppleton
					1148.57	156.25	992.32		
						29.75			
						18.75			

Figure 4: p. 304 of register, see the top and bottom where the pages bend.

Received of the PARSONS PAPER COMPANY, by their Agent, the sum paid to our Signatures, being in full of all demands up to March 1st 1869.

NAMES, 1868	No. POUNDS.	No. REAMS.	No. DAYS.	PRICE.	TOTAL AMOUNT.	RENT, OR BOARD.	BALANCE DUE.	DATE.	SIGNATURES.
Geo. Poppleton			24	18	72.00		72.00		George Poppleton
Halle's Sons			16	18	41.28	7.25	34.00		Halle's Sons
Wm Grant			22	13	51.19		51.19		Wm Grant
Geo A Day			22	15	58.75	7.25	51.50		Geo A Day
S. James			24	15	64.00		64.00		S. James
Chas H Robinson			24	12	48.00		48.00		Chas H Robinson
Ch Magan			24	7	28.58		28.58		Ch Magan
Franc Walker			24	10	42.00		42.00		Franc Walker
Frank Smith			24	9	36.00		36.00		Frank Smith
Anna Casey			23	6	23.75		23.75		Anna Casey
George Whitman			11		11.00	6.00	5.00		George Whitman
Mary Ann			10		10.50	6.25	4.25		Mary Ann
Thomas Burt			24		24.00		24.00		Thomas Burt
John Water			24		24.00	1.20	22.80		John Water
S. Buckland			24		24.00	12.00	12.00		S. Buckland
Emma Carlton			24		24.00		24.00		Emma Carlton
S. Dawson			23		23.50	12.00	11.50		S. Dawson
Adie Dart			22		22.00	12.00	10.00		Adie Dart
Mary Holland			23		23.00		23.00		Mary Holland
Maria Dart			23		23.00	1.20	11.80		Maria Dart
Ellen Hamilton			23		23.00	12.00	11.00		Ellen Hamilton
Adie Colby			22		22.00		22.00		Adie Colby
Mary Burt			24		24.00		24.00		Mary Burt
Frank Water			24		24.00	12.00	12.00		Frank Water
Ella Hill			24		24.00		24.00		Ella Hill
Anna Stone			24		24.00		24.00		Anna Stone
Felia Stockman			24	7	28.00		28.00		Felia Stockman
Maria Brown			22	5	6.60		6.60		Maria Brown
Samy Allen			22	5	48.58	28.23	20.35		Samy Allen
Mary Stone			22	5	44.2	26.11	18.09		Mary Stone
W. A. Ketchum			20	6	28.00	16.24	11.76		W. A. Ketchum
E. S. Grant			20	5	5.99		5.99		E. S. Grant
Maria Brown			25	2	11.40	9.00	2.40		Maria Brown
E. S. Allen			3	90	23.32		23.32		E. S. Allen
Ellen M. Gordon			2	5	13.00		13.00		Ellen M. Gordon
Ellen Pyles			2	11	24.10	12.00	12.10		Ellen Pyles
E. Phillips			2	10	26.20		26.20		E. Phillips
S. Davis			2	18	23.86	12.00	11.86		S. Davis
Martha News			2	16	18.32	12.50	5.82		Martha News
Louise Hill			1	14	22.28		22.28		Louise Hill
Thos Ellen Megar			1	15	3.12		3.12		Thos Ellen Megar
Mrs. Ellen N. Pyle			1	6	7.32		7.32		Mrs. Ellen N. Pyle
Chas. W. News			1	12	11.00		11.00		Chas. W. News
Mary News			1	6	17.26		17.26		Mary News
Frank H. News			1	12	6.25		6.25		Frank H. News
E. Poppleton			1	18	41.28		41.28		E. Poppleton
					1148.57	156.25	992.32		

Figure 5: preview display of p. 198 of register, a red arrow was added to the image indicating where the adjust color interface is

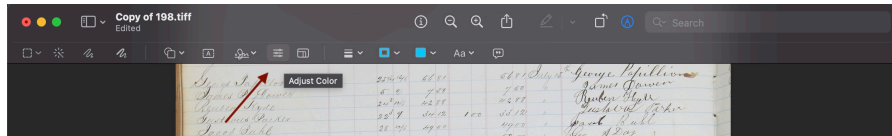


Figure 6: preview display of p. 198 of register after color correcting. In the bottom left is the color adjustment window. The ‘Auto Levels’ button was used for the adjustments.

After rotating, the color correction for the image was opened in preview and the automatic correction was applied, see Figures 5 and 6.

Once all of the edits for the image were made in preview, the image would be renamed to remove the `Copy of` and leave only the number `XXX.tiff`. This indicates that the image has been processed. The process was then repeated to each of the batch of 20 images downloaded locally. Once all of the images were processed in a batch, the download was removed for the batch and another batch of 20 was downloaded and processed. This continued until every image was processed.

2.4 Quality Assurance: Checking for Missing Pages

To identify pages were missing, we created a function that will allow us to see if any file was missing given a sequence of numbers. The function we used is replicated below as the `find_missing_pages()` function.

```
find_missing_pages <- function(start = 0, end = 10,
                               path = "/", suffix = ".png"){
  files_toCheck <- paste0(path, sprintf("%03d", start:end), suffix)
  missing_files <- files_toCheck[!file.exists(files_toCheck)]
  return(missing_files)
}
```

Using this function, we were able to see that beyond the pages that we know are missing, we were missing the scans for several pages. This was really helpful for us as we were able to make another copy of the missing page when we visited the Holyoke Public Library. With the images all converted to `jpeg`s, we were now able to save it into our GitHub repository, and created a interactive Shiny display to help display the images and view the register.

3 Viewing the Register

In addition to the publicly accessible Google Drive (<https://drive.google.com/drive/folders/18Aw57Hhga52E3skMttF7sxDFL6KWO-7q?usp=sharing>) that includes the raw (`.CR2`),

.tiff, and .jpeg files, we have created a web interface to view the scanned pages of the register. This interactive web applet can be found at <https://r.amherst.edu/apps/nhorton/Parsons-Paper>. Each page has its own radio button in the **pages** tab. More details about Parsons Paper Company and the archive are detailed in the **about** tab.

4 Next Steps: Adding to the Digital Commonwealth

We eventually want to submit our results and our project to the Commonwealth of Massachusetts digital archive. Through email communication with the digital commonwealth that a fellow classmate of ours did, we found the required metadata necessary for submission to the archive. They are listed below:

Required:

- file name
- file path
- Type of resource
- BASIC genre/form
- BASIC genre/form: Value URI
- Date type
- Digital Origin
- Host collection
- Physical location (library)
- Rights
- License
- Description standard
- Repository Set Name

We hope to be able work with Dr. Eileen Crosby to submit the scans to the archive by following these guidelines and add them to the Holyoke Library collection.

5 Acknowledgements

We greatly appreciate and would like to thank our instructor (Nicholas Horton), Eileen Crosby (Holyoke Public Library), Tim Pinault (Amherst College), and Zoe Jacobs Feinstein (Amherst College) for their assistance with this project.