

# **STAT210: Parsons Paper Company Payroll Register Curation**

Kevin Jin and Nahia Pino

2023-03-02

## **Table of contents**

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Processes to Curate Ledger</b>	<b>4</b>
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
<b>3</b>	<b>Viewing the Register</b>	<b>10</b>
<b>4</b>	<b>Next Steps: Adding to the Digital Commonwealth</b>	<b>11</b>
<b>5</b>	<b>Acknowledgements</b>	<b>12</b>

## 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](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](https://github.com/STAT210-S23/Parsons_Paper_Register)
- Interactive Web Viewer of Ledger Pages: <https://r.amherst.edu/apps/nhorton/Parsons-Paper>

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 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 1<sup>st</sup> 1867.

NAMES.	No. POUNDS.	No. REAMS.	No. DAYS.	PRICE.	TOTAL AMOUNT.	RENT, OR BOARD.	BALANCE DUE.	DATE.	SIGNATURES.
William Johnson	242 16/6	67 37	27 1/2	67 37	67 37				Wm Johnson
John McEvire	212 4/0	34 95	11 67	30 28					John & Fred
William Connor	25 1/6	6 875	7 25	6 180					William Connor
Martin Kennedy	28 1/0	37 38			37 38				Martin Kennedy
James Casey	28 1/2	51 00			51				James X Casey
Thos. Gallivan	25 0/0	37 50			37 50				Thos. Gallivan
John Vaughan	25 1/4	5 00			5 00				John Vaughan
Pat Doyle	27 9/4	40 60			40 60				Pat Doyle
Sam'l S. Knight	27 8 1/2	55 00			7 28 1/2	47 75			Sam'l S. Knight
Phillip Gilday	28 1/2	57 00			57				Philip Gilday
John Conchen	28 9/4	56 00			56 00	6 25	49 75		John Conchen
Frank Russell	28 1/6	48 12			48 12	9 00	39 12		Frank R. Russell
Charles J. Pepon	28 7/6	35 63			35 63				Wm. J. Pepon
William Donohoy	27 2/1	9 450			9 450				Wm. J. Donohoy
John Flynn	29 1/2	67 50			67 50				John Flynn
E. H. Wellington	29 1/6	87 75			87 75	6 25	81 50		E. H. Wellington
Martin Moran	29 9/0	14 70			14 70				Martin Moran
Wm. McEvire	33 4/6	98 31			98 31	9 33	88 98		Wm. McEvire
Charles Shing	27 1/2	58 50			58 50				Charles Shing
J. A. Allen	27 9/4	40 50			40 50	5 00	35 50		J. A. Allen
Nick Ryan	31 1/4	56 83			56 83				Nick Ryan
Pat & Murphy	31 9/1	46 50			46 50				Pat & Murphy
Rich. Marshall	29 1/4	53 17			53 17				Rich. Marshall
Edward Ward	25 9/0	40 63			40 63				Ed. Ward
Pat & Majors	25 3/6	44 60			44 60	4 67	39 95		Pat & Majors
Pat & O'Kane	27 9/0	48 58			48 58	5 10	38 58		Pat & O'Kane
Joe Beaudear	6 1/0	9 75			9 75	5 00	4 75		Joe Beaudear
Joe Beaudear Jr	25 9/0	48 60			48 60				Joe Beaudear Jr
John Vaughan	27 11/6	11 25			11 25				John Vaughan
Eugene Hanley	26 0/0	39 00			39 00				Eugene Hanley
Liam H. Davis	26 0/0	39 75			39 75				Liam H. Davis
John Shea	18 2/7	27 75			27 75				John Shea
John Warren	3 2/1	75			75				John Warren
Mary Only	11 6/0	12 37			12 37				Mary Only
Kate Nillan	27 9/0	30 37			30 37				Kate Nillan
Kate Moran	3 0/0	27			27				Kate Moran
Kate Baldwin	23 0/0	25 87			25 87				Kate Baldwin
Philip Gilday	7 1/2	7 12			7 12				Philip Gilday
John Flynn	1 00	1 00			1 00				John Flynn
<b>Total</b>	<b>17300 07</b>	<b>6967 16640</b>							
Gustavus Ely	25 1/2	76 50			76 50				Gustavus Ely
Julia Buckley	27 1/2	61 50			61 50				Julia Buckley
J. M. Alden	25 1/2	62 50			62 50				J. M. Alden
Martin Nevel	23 1/6	81 78			81 78	4 00	47 98		Martin Nevel
Pat & Casey	32 2/0	52 82			52 82				Pat & Casey
Austin Ely	27 1/8	81 00			81 00				Austin Ely
Maria Buckley	27 1/8	81 00			81 00				Maria Buckley
J. M. Alden	27 1/2	67 50			67 50				J. M. Alden
Martin Nevel	27 1/6	60 75	4 00		56 75				Martin Nevel
Pat Casey	33 9/4	63 62			53 62				Pat Casey
		343 87	4 00		339 87				
						4 00	3 27		
						3 27	2 00		
						2 00	1 00		

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.

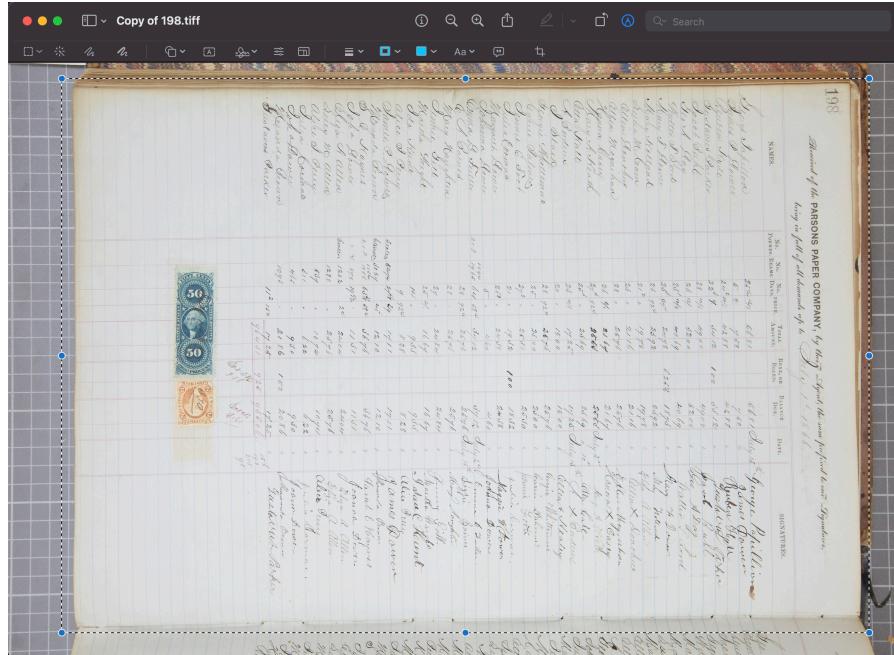


Figure 3: Ex. cropping p. 198 of register in preview

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

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 specified to our Signatures,  
being in full of all demands up to March 1<sup>st</sup> 1869.*

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

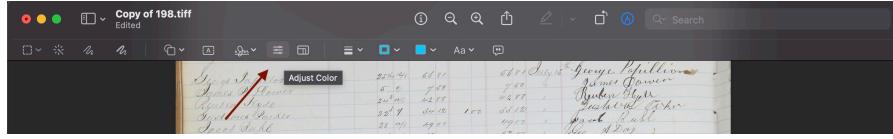


Figure 6: preview display of p. 198 of register after color correcting. In the bottom left is the color adjustment window. The ‘Auto Levels’ button was used for the adjustments.

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 `jpegs`, 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/18Aw57Hgga52E3skMttF7sxDfL6KWO-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.