

# Parsons Paper Company Payroll Register Curation

Kevin Jin and Nahia Pino

2023-02-20

## Introduction

The Parsons Paper Register contains over 300 pages of employee and payroll 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, were scanned on January 10, 2023 as part of this class (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.

This document describes the process of scanning, transforming, and curating the images from the register.

## Conversion Process

Figure 1 displays a sample page from the register. These images were saved as a `.tiff` file that we kept in the `tiff_original` directory in our Google Drive.

As some pages had multiple scans (saved as `XXXa.tiff`, `XXXb.tiff`, 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 of photo, position of page, clarity of the words, and general preference). The rejected scans were moved into a separate folder labeled `deleted`, in the Google Drive.

Within the Desktop Google Drive app, using command selection every photo in the folder was selected and copy-pasted into the same 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.

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 Mc Cune	212 4/0	34 95	11 67	30 28					John & Fred
William Connor	25 1/6	68 75	7 25	61 80					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	50 00			50				John Vaughan
Pat Doyle	27 9/4	40 50			10 50				Pat Doyle
Sam'l S. Knight	27 8 1/2	55 00			47 75				Sam'l S. Knight
Phillip Gilday	28 1/2	57 00			57				Philip Gilday
John Conchen	28 9/4	56 00			49 75				John Conchen
Frank Russell	27 11/6	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 21/2	94 50			94 50				Wm. J. Donohoy
John Flynn	29 1/2	67 50			67 50				John Flynn
E. H. Wellington	27 17/6	57 75	6 25		81 50				E. H. Wellington
Martin Ward	27 9/0	14 70			44 70				Martin Ward
Wm. McCay	33 4/6	98 31	9 33		88 98				Wm. McCay
Charles Shing	27 1/2	58 50			58 50				Charles Shing
J. A. Allen	27 9/4	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. Arkham	29 11/4	53 17			53 17				Rich. Arkham
Edward Ward	25 9/0	40 63			40 63				Ed. Ward
Pat & Majors	25 11/6	44 60	4 67		39 95				Pat & Majors
Pat & O'Kane	27 9/0	48 58	5 10		38 58				Pat & O'Kane
Joe Beaudear	6 1/0	9 75	5 00		4 75				Joe Beaudear
Joe Beaudear Jr	25 9/0	48 60			46 60				Joe Beaudear Jr
John Vaughan	27 11/6	47 25			47 25				John Vaughan
Eugene Hanley	26 0/0	39 00			39				Eugene Hanley
Liam H. Davis	26 0/0	39 75			39 75				Liam H. Davis
John Shea	18 2 1/2	27 75			27 75				John Shea
John Warren	3 2 1/2	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 "	27			27				Kate Moran
Kate Baldwin	23 "	25 87			25 87				Kate Baldwin
Philip Gilday	7 1/2	7 12			7 12				Philip Gilday
John Quinn	1 00	1 00			1 00				John Quinn
<b>Total</b>	<b>173007</b>	<b>6967</b>	<b>16640</b>						
Gustav Ely	25 11/2	76 50			76 50				Gustav Ely
Julia Buckley	27 18/1	61 50			61 50				Julia Buckley
J. M. Alden	25 15/1	62 50			62 50				J. M. Alden
Martin Nevel	23 13/6	81 78	4 00		47 98				Martin Nevel
Pat & Casey	32 2 1/0	52 82			52 82				Pat & Casey
Austin Ely	27 19/	81 00			81 00				Austin Ely
Maria Buckley	27 18/	81 00			81 00				Maria Buckley
J. M. Alden	27 19/	67 50			67 50				J. M. Alden
Martin Nevel	27 19/	60 75	4 00		56 75				Martin Nevel
Pat Casey	33 9/	63 62			53 62				Pat Casey
		343 87	4 00		339 87				
						4 00			
						2 27			
						2 27			
						2 27			
						2 27			

Figure 1: Ex. p. 253 of register - February 1st, 1867

Once the processing folder was set up, the process moved from the online Google Drive platform to the desktop version (which was downloaded). Within the desktop interface, 20 photos were downloaded locally at a time. Starting from the image with the largest number of the downloaded photo, the image was opened with MacOS's `preview` application. In the application, first the photo 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. 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 for example (...).

After cropping, the image was rotated in preview (odds clockwise, evens counterclockwise). After rotating, the color correction for the image was opened in preview and the automatic correction was applied (one button). This was saved.

Once all of the edits were made in preview, the image would be renamed to remove the "Copy of" and leave only the number "XX.tiff". This indicates that the image has been processed. The process was then repeated to each of the batch on 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 photo was processed.

## Check for Missing Pages

To ensure that no 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` method.

```
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)
}
```

From this method, we were able to see that beyond the pages that we know are missing, we were missing the scans for page 368 and 395 (though page 395 was later found). 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.

## **Shiny App**

A Shiny app to display scanned pages 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.