Parsons Paper Company Payroll Register Curation

Kevin Jin and Nahia Pino

2023-02-20

Introduction

The Parsons Paper Copy Register contains near 400pages 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 (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.

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 stamps before. Furthermore, there are some pages where there seems to be scraps of paper attached to them that offer additional context/information about what the page was about. For a current list of pages with these issues, please look at the public Github repository here: https://github.com/STAT210-S23/Parsons_Paper_Register

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

Scanning Process

Nick please finish filling out this part

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 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 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.

Process to convert `.CR2` files to `.tiff`

Within the Desktop Google Drive 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.

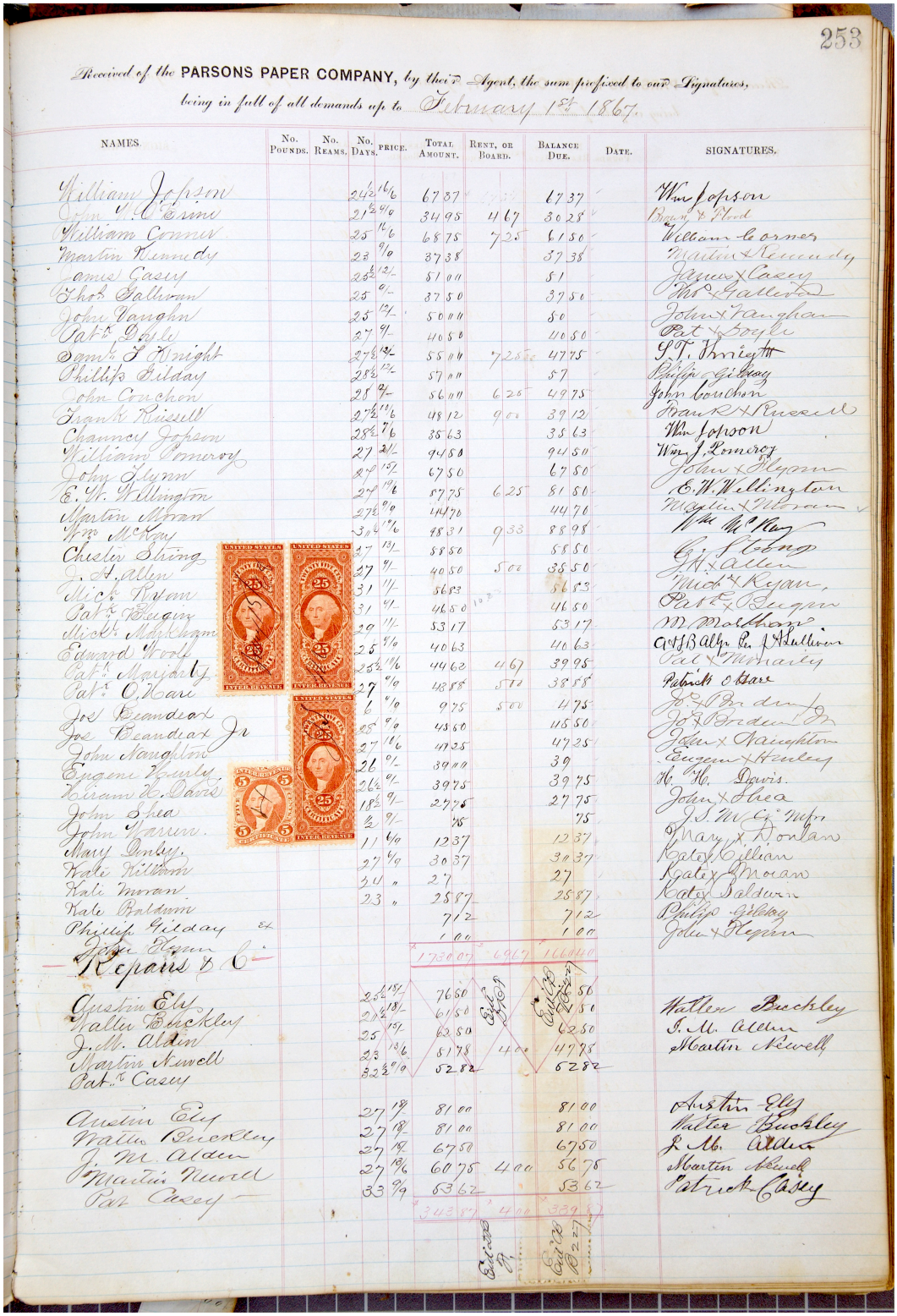
The process moved from the online Google Drive platform to the desktop version (which was downloaded). Within the desktop interface, 20 images were downloaded locally at a time. Starting from the image with the largest number, the image was opened with MacOS's `preview` application. In the application, 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 2 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 3 for an example.

After rotating, the color correction for the image was opened in preview and the automatic correction was applied, see Figures 4 and 5.

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.

Check 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` method.



253

Received of the PARSONS PAPER COMPANY, by their Agent, the sum prefixed 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 Jopson			24½	16/6	67 37		67 37		Wm Jopson
John A Ermi			21½	4/0	34 95	4 67	30 28		Brown & Ford
William Conner			25	14/6	68 75	7 25	61 50		William Conner
Martin Kennedy			23	9/0	37 38		37 38		Martin Kennedy
James Casey			25½	12/-	81 00		81		James Casey
Thos Gallivan			25	9/-	37 50		37 50		Thos Gallivan
John Vaughn			25	12/-	50 00		80		John Vaughan
Pat.k Doyle			27	9/-	40 50		40 50		Pat Doyle
Sam.l T Knight			27½	12/-	55 00	7 25	47 75		S T. Knight
Phillips Gilday			28½	12/-	57 00		57		Phillips Gilday
John Couchon			28	9/-	56 00	6 25	49 75		John Couchon
Frank Russell			27½	11/6	48 12	9 00	39 12		Frank Russell
Chauncy Jopson			28½	7/6	35 63		35 63		Wm Jopson
William Comeroy			27	21/-	94 50		94 50		Wm J Comeroy
John Glynn			27	15/-	67 50		67 50		John Glynn
E. W. Willington			27	12/6	87 75	6 25	81 50		E W Willington
Martin Moran			27½	9/0	44 70		44 70		Martin Moran
Wm McKay			30¼	12/6	48 31	9 33	38 98		Wm McKay
Chester Strong			27	13/-	38 50		38 50		C Strong
J. A. Allen			27	9/-	40 50	5 00	35 50		J A Allen
Mich.l Ryan			31	11/-	56 83		56 83		Mich Ryan
Pat.k Eugene			31	9/-	46 50		46 50		Pat Eugene
Mich.l Markham			29	11/-	53 17		53 17		M Markham
Edward Wolf			25	9/0	40 63		40 63		Wolf & Ally Pr. J Sullivan
Pat.k Murphy			25½	11/6	44 62	4 67	39 95		Pat Murphy
Pat.k O'Hare			27	9/0	48 88	5 00	38 88		Patrick O'Hare
Jos Beandear			6	9/0	9 75	5 00	4 75		Jos Beandear
Jos Beandear Jr			25	9/0	45 60		45 50		Jos Beandear Jr
John Naughton			27	17/6	47 25		47 25		John Naughton
Eugene Tourly			26	9/-	39 00		39		Eugene Tourly
Hiram C. Davis			26½	9/-	39 75		39 75		H. H. Davis.
John Shea			18½	9/-	27 75		27 75		John Shea
John Sarum			½	9/-	75		75		J S M & Co Mfrs
Mary Dinley			11	6/0	12 37		12 37		Mary Dinley
Kate Killam			27	4/0	30 37		30 37		Kate Killam
Kate Moran			24	"	27		27		Kate Moran
Kate Baldwin			23	"	25 87		25 87		Kate Baldwin
Philip Gilday					7 12		7 12		Philip Gilday
John Eagan					1 00		1 00		John Eagan
Repairs &c					1730 00	69 47	1660 40		
Austin Ely			25½	18/-	76 50		76 50		
Walter Buckley			20½	18/-	61 50		61 50		Walter Buckley
J. M. Allen			25	15/-	62 50		62 50		J M Allen
Martin Newell			23	18/6	81 78	4 00	47 78		Martin Newell
Pat.k Casey			32½	9/0	52 82		62 82		
Austin Ely			27	18/	81 00		81 00		Austin Ely
Walter Buckley			27	18/	81 00		81 00		Walter Buckley
J. M. Allen			27	15/	67 50		67 50		J M Allen
Martin Newell			27	19/6	60 75	4 00	56 75		Martin Newell
Pat Casey			33	9/9	63 62		53 62		Patrick Casey
					343 87	4 00	339 87		

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

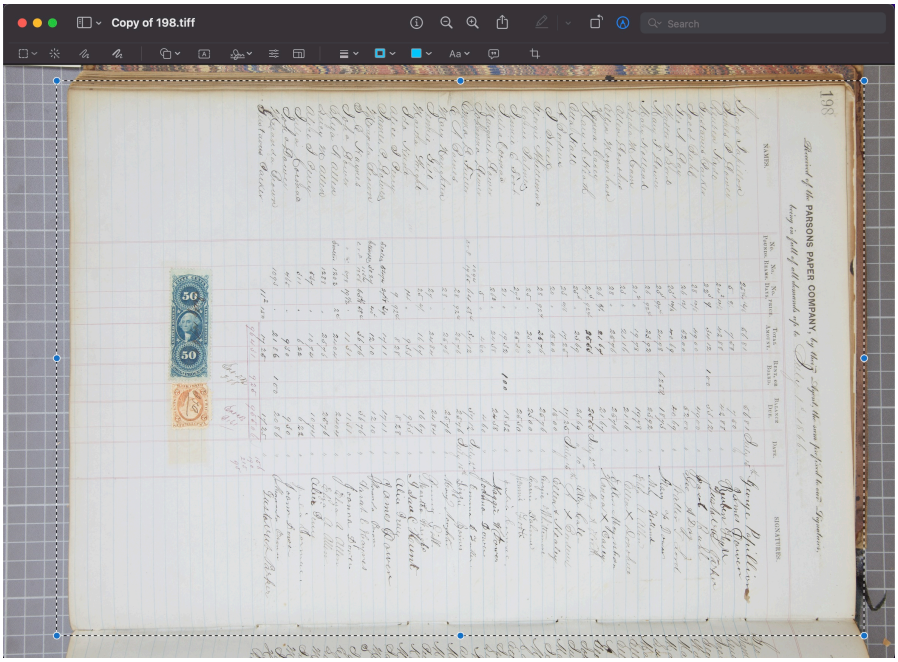


Figure 2: Ex. cropping p. 198 of register in preview

```
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 `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.

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/

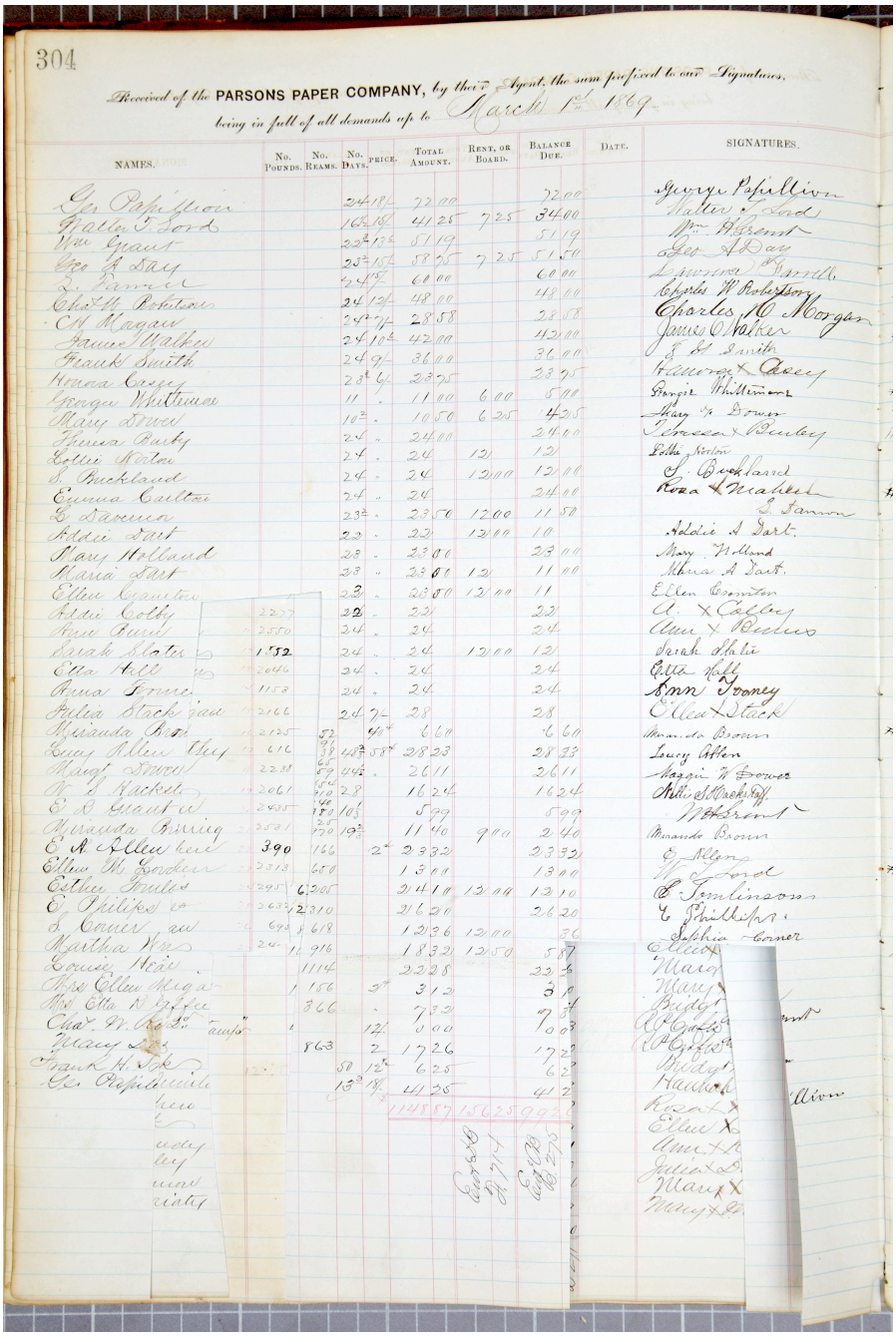


Figure 3: p. 304 of register, see the top and bottom where the pages bend.

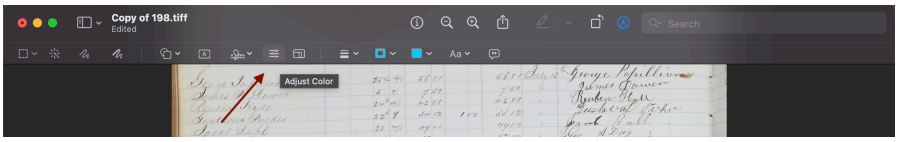


Figure 4: preview display of p. 198 of register, a red arrow was added to the image indicating where the adjust color interface is

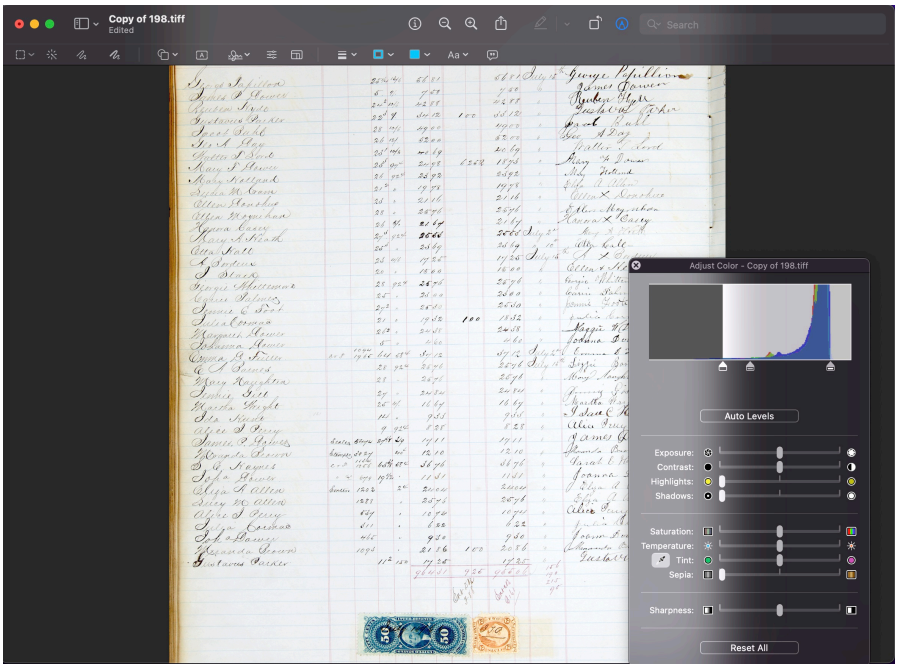


Figure 5: preview display of p. 198 of register after color correcting. In the bottom left is the colord adjustment window. The 'Auto Levels' button was used for the adjustments.

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.

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 to submit our collection to the archive by following these guidelines.

Acknowledgements

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