

STAT210: Parsons Paper Company Payroll Register Curation

Kevin Jin and Nahia Pino

2023-05-21

Table of contents

1	Introduction	2
2	Processes to Curate Ledger	5
2.1	Scanning Process	5
2.2	Conversion Process	6
2.3	Process to convert .CR2 files to .tiff	8
2.4	Quality Assurance: Checking for Missing Pages	10
3	Viewing the Register	11
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. An example page, after conversion to jpeg and color correction, is shown in Figure 1. This could give insight on the daily lives of the workers for the biggest company of the time period.

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 2).

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 specified to our Signatures,
being in full of all demands up to January 1861.

NAMES.	NO. POUNDS.	NO. REAMS.	NO. DAYS.	PRICE.	TOTAL AMOUNT.	RENT, OR BOARD.	BALANCE DUE.	DATE.	SIGNATURES.
Ellen Healy	160	21	50	15	15.00		15.00	Jan 1st	Ellen & Healy
Anna Clark	100	21	4	12	12.00		12.00	Jan 6	Anna Clark
C. Atwood	125	17	4	8	8.00		8.00	"	C. Atwood
C. Atwood	160	22	4	11	11.00		11.00	"	C. Atwood
C. Harrington	172	20	4	11	11.00		11.00	"	C. Harrington
Catharine Gerringer	180	21	4	12	12.00		12.00	"	C. Harrington
Mary Shedd	172	20	4	12	12.00		12.00	"	M. Shedd
Mary Swartz	172	20	4	11	11.00		11.00	"	Mary & Swartz
Chamala G. Ford	180	21	4	12	12.00		12.00	"	Chamala G. Ford
Anna Shedd	180	21	4	12	12.00		12.00	"	Anna Shedd
Souisa Emerson	172	20	4	11	11.00		11.00	"	Souisa Emerson
Ellen Sawyer	180	21	4	12	12.00		12.00	"	Ellen Sawyer
Souisa S. S. S.	172	20	4	11	11.00		11.00	"	Souisa S. S. S.
Mary Campbell	180	21	4	12	12.00		12.00	"	Mary & Campbell
Catharine Guly	172	20	4	11	11.00		11.00	"	Catharine Guly
Mary Ann	160	22	4	11	11.00		11.00	"	Mary Ann
Mary Donohue	180	21	4	12	12.00		12.00	"	Mary & Donohue
Mrs. Gove	172	20	4	11	11.00		11.00	"	Mrs. Gove
Mrs. W. J. Barnes	160	21	4	10	10.00		10.00	"	Mrs. W. J. Barnes
Mary Taft	172	20	4	11	11.00		11.00	"	Mary Taft
Ellie Connor	160	18	"	9	9.00		9.00	"	Ellie Connor
S. H. Hobart	160	20	100	85	85.00		85.00	"	S. H. Hobart
Mrs. Dowd	20	84	"	15	15.00	4.00	11.00	"	Mrs. Dowd
Emily Bowditch	24	80	"	12	12.00		12.00	"	Emily Bowditch
Margaret Haffy	20	11	"	12	12		12	"	Margaret Haffy
					\$18.62	11.00	50.00		
<u>Finishes</u>									
Geo. H. Perry	28	300	65	0.00			87.00	Jan 10th	Geo. H. Perry
John Dowd	20	8	38	0.00	7.00		21.00	"	John Dowd
St. J. Money	20	76	60	0.12			30.60	"	George J. Money
C. Hatfield	22	46	26	0.18			20.38	"	C. Hatfield
Edwin M. Wilson	20	7	50	0.15			30.15	"	Ed. M. Wilson
Charles Elly	21	74	15	0.15			30.15	"	Charles Elly
H. A. Hyde	20	74	30	0.12			30.00	Aug 12th	H. A. Hyde
P. F. Greenport	13	93	17	0.17			17.57	Jan 15th	P. F. Greenport
John Dowd	20	4	12	0.10			20.00	"	John Dowd
Mary Dowd	18	1	9	0.17			9.07	"	Mary Dowd
Catharine Dowd	19	1	11	0.17			11.07	"	Catharine Dowd
Margaret Dowd	19	1	11	0.17			11.07	"	Margaret Dowd
Emma Perry	11	5	7	0.11			7.07	"	Emma Perry
Mary Melody	18	1	11	0.11			11.01	"	Mary Melody
Silvia A. Smith	12	2	7	0.12			7.02	"	Silvia A. Smith
Mary Melody	20	7	14	0.09			14.09	"	Mary Melody
K. A. Burnham	14	4	8	0.15			8.05	"	K. A. Burnham
Ellen Perry	11	1	6	0.11			6.07	"	Ellen Perry
Lucy Allen	20	6	15	0.12			15.12	"	Lucy Allen
Silvia M. Evans	10	3	0	12.00			12.00	"	Silvia M. Evans
Erinda Perry	1	10	1	6.12			6.12	"	Erinda Perry
S. J. Money	1	10	1	1.28			1.28	"	S. J. Money
A. J. Watford	66	89	15	0.00	7.50		7.50	"	A. J. Watford
					\$78.59	11.00	57.09		

Figure 1: page 002 of register after color correcting. This what a typical page of the registry looks like, including indications of thier jobs and location.

Figure 2: page 286 of register after color correcting. On the right 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 3 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 the publicly accessible 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,

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	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	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	7 28	47 75					G. T. Knight
Phillip Gilday	28 1/2	57 00			57				Philip Gilday
John Conchen	28 9/4	56 00	6 25	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	51 50					E. H. Wellington
Martin Ward	27 9/0	14 70			14 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 1/4	53 17			53 17				Rich. Arkham
Edward Ward	25 9/0	40 63			40 63				Ed. Ward
Pat & Majors	25 3/6	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	17 25			17 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
Total	17300 00	6967 16640							
Austin Ely	25 1/2	76 50			76 50				Austin Ely
Julia Buckley	27 18/1	61 50			61 50				Julia Buckley
J. M. Alden	25 1/2	62 50			62 50				J. M. Alden
Martin Nevel	23 13/6	51 78	4 00	47 78					Martin Nevel
Pat Casey	32 2 1/0	52 82			52 82				Pat Casey
Austin Ely	27 1/8	81 00			81 00				Austin Ely
Julia Buckley	27 18/1	81 00			81 00				Julia 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 9/9	63 62			53 62				Pat Casey
		343 87	4 00	339 87					
		<i>Given</i>	<i>Given</i>	<i>Given</i>					

Figure 3: Ex. page 253 of register - part of the February 1st, 1867 payroll

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.

2.3 Process to convert .CR2 files to .tiff

Within the Google Drive website, 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.

After creating all of the copies, this folder was opened on the 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 4 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 5 for an example.

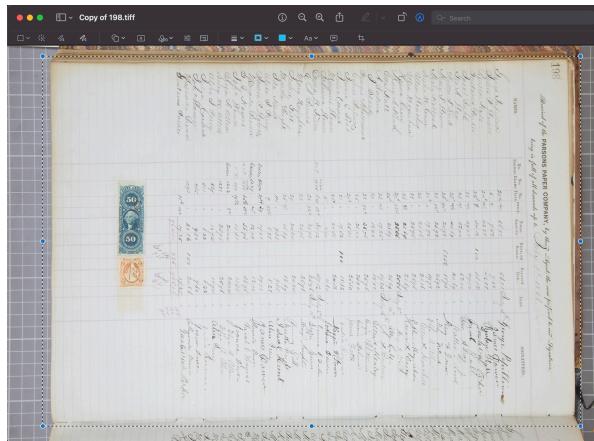


Figure 4: Example cropping of page 198 of register in preview in order to keep all of the edges in view.

After rotating, the color correction for the image was opened in preview and the automatic correction was applied, see Figures 6 and 7.

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

Figure 5: page 304 of register, see the top and bottom where the pages bend.

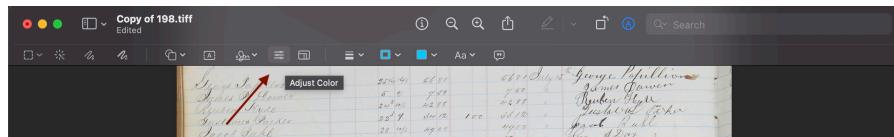


Figure 6: preview display of page 198 of register, a red arrow was added to the image indicating where the adjust color interface is

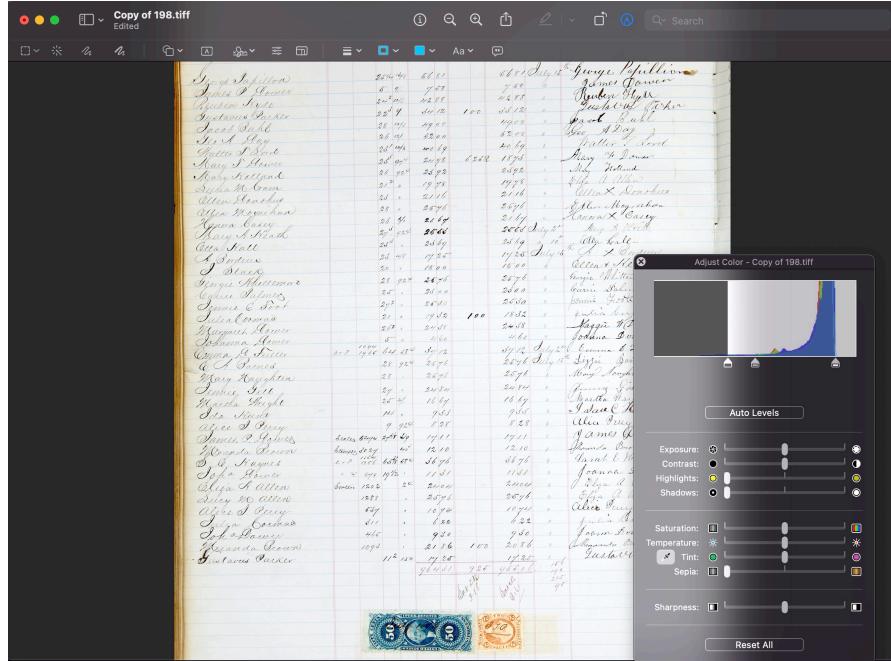


Figure 7: preview display of p. 198 of register after color correcting. In the bottom left is the colord adjustment window. The ‘Auto Levels’ button was used for the adjustments.

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, which was written to automate this process.

```
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 scan 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/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.