Extracting Data From PDFs

SIMPLE STRATEGIES TO LIBERATE TRAPPED DATA

EMILY LE COZ | elecoz@usatoday.com

IRE 2023 ORLANDO

SIGN UP FOR GOOGLE PINPOINT + THE BETA DATA EXTRACTION TOOL



Sign into Google Pinpoint here:

journaliststudio.google.com/pinpoint/about

Wait for approval (usually very quick)

Request access to to the beta data extraction tool here: shorturl.at/xABU4

Wait for approval (also relatively quick)

WHAT'S A PDF?

PORTABLE DOCUMENT FORMAT

Created in 1993 by Adobe, PDFs allowed us to "capture" documents from any application, send them electronically, and be able to open, view and print them on any machine.

GOOD FOR ...

Viewing information in its original format

BAD FOR ...

Sorting, analyzing or working with the information

WHY PUT DATA IN A PDF THEN?

So nobody can sort, analyze or work with the data!

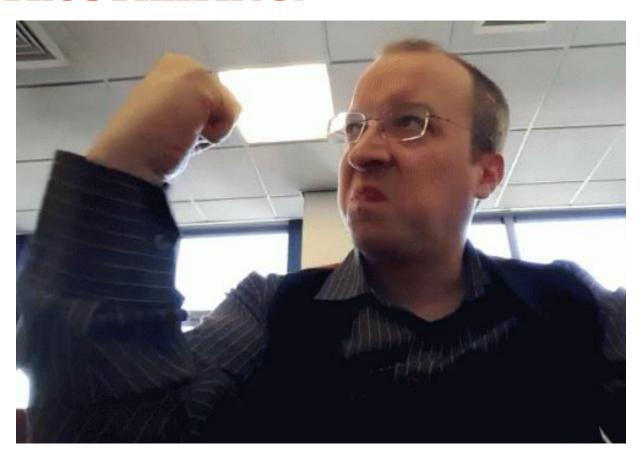
Government agencies that are loathe to share their data can provide it to you in the worst possible format while still complying with applicable public records laws.

Scanned (image-based)
PDFs

Non-tabular data PDFs

Hand-written data PDFs

VERY FRUSTRATING!





LET'S GET TO WORK



EASY CASE SCENARIO

Tabular data inside a text-based PDF

Massachusetts Department of Public Health opioid overdose death data

DOWNLOAD: https://www.mass.gov/lists/current-opioid-statistics

- → Updated data as of December 2022
- → Opioid overdose deaths by city/town
- → Download PDF



Number of Opioid-Related Overdose Deaths, All Intents by City/Town 2015-2021

Massachusetts Department of Public Health

POSTED: DECEMBER 2022

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the "Death Occurrence" table (Table 2) as "Outside of Massachusetts". Non-Massachusetts residents who died in Massachusetts are not included in the "Residence" table (Table 1).

Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2021. For 2017 to 2021, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly and all new confirmed cases will be included in the table below with previously confirmed cases.

City/Town of				Year of Death			
Residence	2015	2016	2017	2018	2019	2020	2021
Abington	6	3	4	11	4	2	9
Acton	4	7	1	4	0	2	4
Acushnet	4	4	7	2	1	2	2
Adams	3	1	0	7	2	1	5
Agawam	5	4	8	12	13	17	8
Alford	0	0	0	0	0	0	0
Amesbury	7	8	10	9	7	5	8
Amherst	1	2	3	2	3	2	
Andover	6	3	2	2	6	0	8
Aguinnah	1	0	0	0	0	0	0
Arlington	6	6	3	3	4	4	3

HOW DO YOU KNOW IT'S TEXT-BASED?

Find Acton

Previous

Next

Table 2. Number of confirmed opioid-related overdose deaths for all intents by city/town of death occurrence, 2015-2021. For 2019 to 2021, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly and all new confirmed cases will be included in the table below with previously confirmed cases.

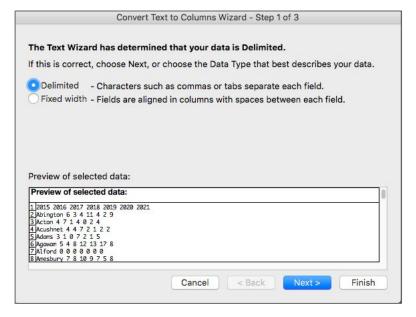
City/Town of				Year of Dea	th		
Death Occurrence	2015	2016	2017	2018	2019	2020	2021
Abington	6	2	3	5	4	1	4
Acton	2	2	0	1	0	3	3
Acushnet	4	2	4	0	0	1	0
Adams	3	1	0	4	1	2	2
Agawam	2	0	4	8	4	8	4
Alford	0	0	0	0	0	0	0
Amesbury	6	1	4	6	2	2	7
Amherst	1	3	1	2	1	0	0
Andover	2	3	2	2	0	0	5
Aguinnah	0	0	0	0	0	0	0
Arlington	4	2	1	1	5	2	1

CAN I JUST COPY-PASTE IT INTO EXCEL? YES, BUT ...

A	В	C	D	E	F
2015 2016 20	17 2018 2019	2020 2021			
Abington 6 3	4 11 4 2 9				
Acton 4 7 1 4	024				
Acushnet 4 4	72122				
Adams 3 1 0 7	7215	L.			
Agawam 5 4 8	3 12 13 17 8				
Alford 0 0 0 0	000				
Amesbury 7 8	109758				
Amherst 1 2 3	32321				
Andover 6 3 2	2608				
Aquinnah 10	00000				
Arlington 6 6	3 3 4 4 3				
Ashburnham	1201220				
Ashby 0 0 0 0	112				
Ashfield 100	1000				
Ashland 4 4 4	2311				
Athol 3 3 6 6	7 14 12	I.			
Attleboro 10	18 26 11 21 10	22			
		75.77			

USE TEXT-TO-COLUMNS FUNCTION





Delimete	ers						
Tab						Treat consecutive delimiters as one	
Semio	colon					Text qualifier:	
Comn	ma					and the state of t	
Space	e						
Other							
	:	cted	l dat	a:			
Other	:	cted	l dat	a:			
Other Preview	of sele			5.4	2021		
Other	of sele			5.4	2	3 1	7
Other Preview	of sele	2018	2019 11 4	2020	2) 	
Other Preview	of sele			5.4	2	2 2 5 6 8	

FORMAT COLUMN HEADERS & TADA!*

A	В	С	D	E	F	G	Н
	2015	2016	2017	2018	2019	2020	2021
Abington	6	3	4	11	4	2	9
Acton	4	7	1	4	0	2	4
Acushnet	4	4	7	2	1	2	2
Adams	3	1	0	7	2	1	5
Agawam	5	4	8	12	13	17	8
Alford	0	0	0	0	0	0	0
Amesbury	7	8	10	9	7	5	8
Amherst	1	2	3	2	3	2	1
Andover	6	3	2	2	6	0	8
Aquinnah	1	0	0	0	0	0	0
Arlington	6	6	3	3	4	4	3
Ashburnham	1	2	0	1	2	2	0
Ashby	0	0	0	0	1	1	2
Ashfield	1	0	0	1	0	0	0
Ashland	4	4	4	2	3	1	1
Athol	3	3	6	6	7	14	12
Attleboro	10	18	26	11	21	10	22

*BUT DO YOU
REALLY WANT
TO REPEAT
THIS PROCESS
FOR ALL 16
PAGES?

NO.

EXTRACT DATA WITH TABULA



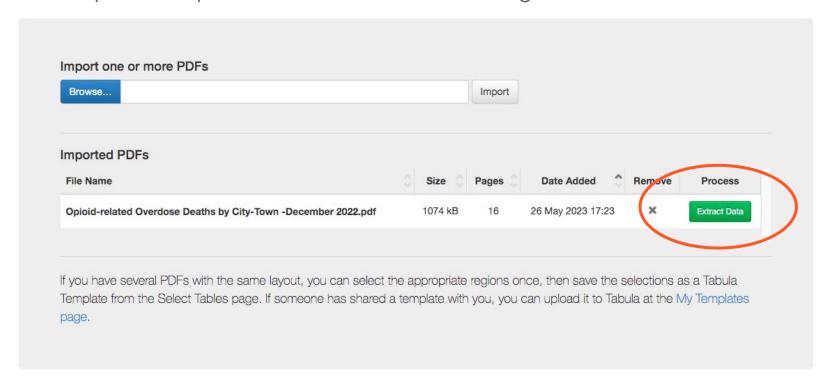
Tabula is a free extraction tool created by journalists to turn tabular data locked inside text-based PDFs (not scanned/image-based) into exportable CSV or spreadsheet files.

DOWNLOAD: https://tabula.technology/

- → Select Windows, Mac or Github version and download
- → Extract the ZIP file and launch the Tabula program
- → Open the program from the menu bar

IMPORT YOUR FILE

Let's import the opioid deaths PDF. Then hit the green "extract data" button.





Number of Opioid-Related Overdose Deaths, All Intents by City/Town 2015-2021

Massachusetts Department of Public Health

POSTED: DECEMBER 2022

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the "Death Occurrence" table (Table 2) as "Outside of Massachusetts". Non-Massachusetts residents who died in Massachusetts are not included in the "Residence" table (Table 1).

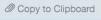
Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2021. For 2017 to 2021, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly and all new confirmed cases will be included in the table below with previously confirmed cases.

City/Town of				Year of Death	i.		
Residence	2015	2016	2017	2018	2019	2020	2021
Abington	6	3	4	11	4	2	9
Acton	4	7	1	4	0	2	4
Acushnet	4	4	7	2	1	2	2
Adams	3	1	0	7	2	1	5
Agawam	5	4	8	12	13	17	8
Alford	0	0	0	0	0	0	0
Amesbury	7	8	10	9	7	5	8
Amhoret	1	2	3	2	3	2	1



Export Format: CSV

Export



Preview of Extracted Tabular Data

City/Town of				Year of Death			
Residence	2015	2016	2017	2018	2019	2020	2021
Abington	6	3	4	11	4	2	9
Acton	4	7	1	4	0	2	4
Acushnet	4	4	7	2	1	2	2
Adams	3	1	0	7	2	1	5
Agawam	5	4	8	12	13	17	8
Alford	0	0	0	0	0	0	0
Amesbury	7	8	10	9	7	5	8
Amherst	1	2	3	2	3	2	1
Andover	6	3	2	2	6	0	8
Aquinnah	1	0	0	0	0	0	0
Arlington	6	6	3	3	4	4	3
Ashburnham	1	2	0	1	2	2	0
A - L L	0	0		0			



Α	В	C	D	E	F	G	H	
City/Town of				Year of Death				
Residence	2015	2016	2017	2018	2019	2020	2021	
Abington	6	3	4	11	4	2	9	
Acton	4	7	1	4	0	2	4	
Acushnet	4	4	7	2	1	2	2	
Adams	3	1	0	7	2	1	5	
Agawam	5	4	8	12	13	17	8	
Alford	0	0	0	0	0	0	0	
Amesbury	7	8	10	9	7	5	8	
Amherst	1	2	3	2	3	2	1	
Andover	6	3	2	2	6	0	8	
Aquinnah	1	0	0	0	0	0	0	
Arlington	6	6	3	3	4	4	3	
Ashburnham	1	2	0	1	2	2	0	
Ashby	0	0	0	0	1	1	2	
Ashfield	1	0	0	1	0	0	0	
Ashland	4	4	4	2	3	1	1	
Athol	3	3	6	6	7	14	12	
Attleboro	10	18	26	11	21	10	22	
Auburn	4	2	4	6	4	3	4	
Avon	3	3	2	2	1	3	3	
Ayer	1	8	1	2	4	1	1	
Barnstable	13	21	13	18	18	14	15	
Barre	3	0	0	3	3	1	0	
Becket	0	0	0	2	0	0	0	
Bedford	1	6	3	3	4	3	3	
Belchertown	2	2	5	7	3	4	1	
Bellingham	3	5	4	8	4	2	6	
Belmont	4	4	4	2	2	1	2	

HARDER CASE SCENARIO

Tabular data inside an image-based PDF

City of Regina, Saskatchewan, Wastewater Treatment Plant Monthly Report

DOWNLOAD: https://shorturl.at/dik06

- → Click on July 2022 report
- → Download PDF

		July 20						luent Resu			ent Pl		0050853-0	7-00	
Date	Flow to Creek	Total Coliform	E. coli	Enterococci	Dissolved Oxygen (Field Test)	Temp. (Field Test)	Chloride (Field Test)	Total Phosphorus	рН @15±1°C	Total Suspended Solids	CBOD	Ammonia Nitrogen	Unionized Ammonia @15±1°C	Total Kjeldahl Nitrogen	Total Nitrogen
	(ML)		(MPN /100mL)		(mg/L)	(°C)	(mg/L)	(mg P/L)	(SU)	(mg/L)	(mg/L)	(mg N/L)	(mg N/L)	(mg N/L)	(mg N/L)
01	56.51	178	18		7.35	17.7		0.31		6.0		0.60		2.4	9.3
02	56.53		3.00		7.30	17.6		0.36		8.2		1.20		3.2	9.
03	58.81				7.37	17.8	190	0.35	7.49	8.6	3.4	0.88	< 0.030	2.7	9.5
04	64.02	461	34		6.52	17.6	196	0.54	7.48	12.0	3.9	1.20	< 0.030	3.5	10.7
05	68.67	866	47	78	7.15	17.6	185	0.89	7.42	22.6	6.8	1.96	< 0.030	5.1	11.6
06	60.69	1,840	260		7.75	17.8	183	0.92	7.50	23.2	6.5	1.94	< 0.030	5.3	12.4
07	52.52	1,300	98		8.28	17.8	180	0.98	7.48	24.6	6.9	1.06	< 0.030	4.5	11.5
08	61.96	1,630	161		8.94	18.2		0.69		15.4		0.93		4.0	11.6
09	58.16				7.81	18.5		0.58		12.9		0.52		2.8	10.6
10	58.74				8.17	18.3	180	0.50	7.51	10.8	4.3	0.73	< 0.030	3.0	11.6
11	60.83	461	51	33	8.07	18.1	184	0.60	7.49	13.4	5.0	0.40	< 0.030	3.1	9.1
12	59.75	727	69		7.43	18.4	188	0.38	7.48	7.4	4.1	1.02	< 0.030	3.2	8.4
13	59.75	821	161		7.27	18.7	195	0.91	7.54	26.8	6.1	1.08	< 0.030	4.6	9.1
14	54.90	1,540	56		5.22	18.7	197	0.62	7.44	16.4	5.0	2.50	< 0.030	5.5	10.3
15	61.28	3,970	118		6.12	18.9		0.61		9.4		1.32		4.1	9.3
16	58.48	0.000.000			7.94	19.0		0.47		13.4		0.80		3.4	7.6
17	59.70				8.18	19.1	189	0.63	7.52	18.0	5.3	1.40	< 0.030	4.6	8.5
18	56.68	172	12	15	7.13	18.9	182	0.90	7.30	27.2	8.5	4.12	< 0.030	7.1	10.5
19	133.45	6,870	921		7.28	17.6	102	0.38	7.39	10.2	4.9	0.90	< 0.030	3.2	7.6
20	90.21	102	10		9.31	17.0	129	0.34	7.53	9.2	3.6	0.47	< 0.030	2.6	8.7
21	81.68	194	4		8.87	17.8	175	0.24	7.49	6.2	2.6	2.07	< 0.030	4.3	8.9
22	68.61	613	65		8.55	18.1	25500	0.13	00000	2.8	3000	2.43		4.1	8.6
23	64.36				8.72	18.2		0.11		2.7		1.81		3.3	8.1
24	64.77				8.60	18.6	195	0.12	7.43	2.6	< 2.0	2.03	< 0.030	3.8	9.5
25	65.36	488	16		7.62	18.6	200	0.13	7.51	4.0	2.4	1.81	< 0.030	3.6	9.0
26	70.92	649	43	23	7.88	18.9	196	0.13	7.52	3.4	2.3	1.71	< 0.030	3.4	7.0
27	80.53	109	4		7.27	19.0	195	0.11	7.45	3.2	< 2.0	1.92	< 0.030	3.5	7.6
28	75.96	47	2		7.31	19.2	194	0.12	7.57	2.9	< 2.0	1.58	< 0.030	3.4	8.3
29	73.52	121	5		6.69	19.3		0.13		3.4		0.83		2.6	7.2
30	72.37				6.84	19.5		0.13		3.3		1.19		3.0	7.5
31	70.69				8.13	19.7	197	0.12	7.49	4.0	2.1	1.41	< 0.030	3.3	7.6
um	2080.41										,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,				
Avg	67.11				7.65	18.4	182	0.43	7.48	11	4	1	< 0.030	3.7	9
Vlin	52.52				5.22	17.0	102	0.11	7.30	2.6	< 2.0	0.40	< 0.030	2.4	7.0
Max	133.45		921		9.31	19.7	200	0.98	7.57	27.2	8.5	4.12	< 0.030	7.1	12.4
Pern	nit Limits		≤1000					≤0.75		≤15	≤15	≤4	≤1.24	10	≤ 10

HOW DO YOU KNOW IT'S IMAGE-BASED?



LET'S OCR IT WITH A FREE ONLINE TOOL





ilovepdf.com

tools.pdf24.org/en







JPG to PDF

PDF to Word



ilovepdf.com



Merge PDF

Combine PDFs in the order you want with the easiest PDF merger available.



Split PDF

Separate one page or a whole set for easy conversion into independent PDF files.



Compress PDF

Reduce file size while optimizing for maximal PDF quality.



PDF to Word

Easily convert your PDF files into easy to edit DOC and DOCX documents. The converted WORD document is almost 100% accurate.



PDF to Powerpoint

Turn your PDF files into easy to edit PPT and PPTX slideshows.



PDF to Excel

Pull data straight from PDFs into Excel spreadsheets in a few short seconds.



Word to PDF

Make DOC and DOCX files easy to read by converting them to PDF.



Powerpoint to PDF

Make PPT and PPTX slideshows easy to view by converting them to PDF.



Excel to PDF

Make EXCEL spreadsheets easy to read by converting them to PDF



Edit PDF

Add text, images, shapes or freehand annotations to a PDF document. Edit the size, font, and color of the added content.



PDF to JPG

Convert each PDF page into a JPG or extract all images contained in a PDF.



JPG to PDF

Convert JPG images to PDF in seconds, Easily adjust orientation and margins.



Sign PDF

Sign yourself or request electronic signatures from others.



Watermark

Stamp an image or text over your PDF in seconds. Choose the typography, transparency and position.



Rotate PDF

Rotate your PDFs the way you need them. You can even rotate multiple PDFs at once!



HTML to PDF

Convert webpages in HTML to PDF. Copy and paste the URL of the page you want and convert it to PDF with a click.



Unlock PDF

Remove PDF password security, giving you the freedom to use your PDFs as you want.



Protect PDF

Protect PDF files with a password. Encrypt PDF documents to prevent unauthorized access.



Organize PDF

Sort pages of your PDF file however you like. Delete PDF pages or add PDF pages to your document at your convenience.



PDF to PDF/A

Transform your PDF to PDF/A, the ISOstandardized version of PDF for long-term archiving. Your PDF will preserve formatting when accessed in the future.



Repair PDF

Repair a damaged PDF and recover data from corrupt PDF. Fix PDF files with our Repair tool.



Page numbers

Add page numbers into PDFs with ease. Choose your positions, dimensions, typography.



Scan to PDF

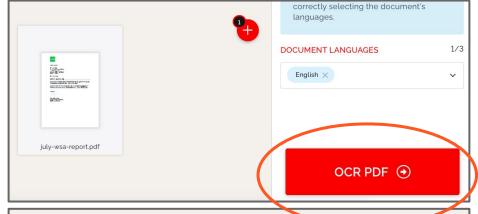
Capture document scans from your mobile device and send them instantly to your browser.



OCR PDF

Easily convert scanned PDF into searchable and selectable documents. 







1

Choose files

2

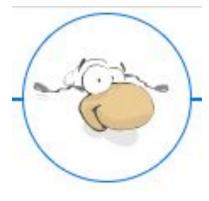
Language English ▼

3

Start OCR

4



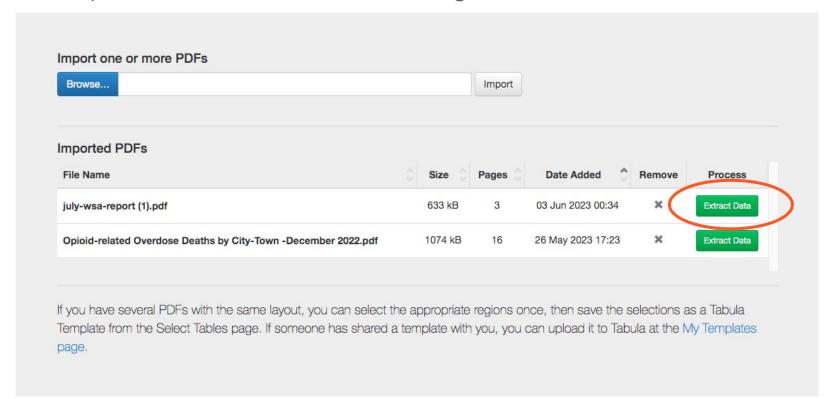


LET'S CHECK THE OCR JOB...

			0.										rind		
			CI	ty of I	Kegir	ia - N	wast	ewa	ter I	reatm	ent Pl	ant	Wastew	ater	•
		July 2	022	0.	z 2		Final Effl	uent Resi	ults			Permit# 0	0050853-07	Previous	Next
Date	Flow to Creek	Total Coliform	E. coli	Enterococci	Dissolved Oxygen (Field Test)	Temp. (Field Test)	Chloride (Field Test)	Total Phosphorus	pH @15±1°C	Total Suspended Solids	CBOD	Ammonia Nitrogen	Unionized Ammonla @15±1°C	Total Kjeldahl Nitrogen	Total Nitrogen
	(ML)		(MPN /100mL)		(mg/L)	(°C)	(mg/L)	(mg P/L)	(SU)	(mg/L)	(mg/L)	(mg N/L)	(mg N/L)	(mg N/L)	(mg N/L)
01	56.51	178	18		7.35	17.7		0.31		6.0		0.60		2.4	9.3
02	56.53				7.30	17.6		0.36		8.2		1.20		3.2	9.2
03	58.81				7.37	17.8	190	0.35	7.49	8.6	3.4	0.88	< 0.030	2.7	9.9
04	64.02	461	34		6.52	17.6	196	0.54	7.48	12.0	3.9	1.20	< 0.030	3.5	10.7
05	68.67	866	47	78	7.15	17.6	185	0.89	7.42	22.6	6.8	1.96	< 0.030	5.1	11.6
06	60.69	1,840	260		7.75	17.8	183	0.92	7.50	23.2	6.5	1.94	< 0.030	5.3	12.4
07	52.52	1,300	98		8.28	17.8	180	0.98	7.48	24.6	6.9	1.06	< 0.030	4.5	11.5
08	61.96	1,630	161		8.94	18.2		0.69		15.4		0.93		4.0	11.6
09	58.16				7.81	18.5		0.58		12.9		0.52		2.8	10.6
10	58.74				8.17	18.3	180	0.50	7.51	10.8	4.3	0.73	< 0.030	3.0	11.6
11	60.83	461	51	33	8.07	18.1	184	0.60	7.49	13.4	5.0	0.40	< 0.030	3.1	9.1
12	59.75	727	69		7.43	18.4	188	0.38	7.48	7.4	4.1	1.02	< 0.030	3.2	8.4
13	59.75	821	161		7.27	18.7	195	0.91	7.54	26.8	6.1	1.08	< 0.030	4.6	9.1
14	54.90	1,540	56		5.22	18.7	197	0.62	7.44	16.4	5.0	2.50	< 0.030	5.5	10.3
15	61.28	3,970	118		6.12	18.9		0.61		9.4		1.32		4.1	9.3
16	58.48		100000		7.94	19.0		0.47		13.4		0.80		3.4	7.6
17	59.70				8.18	19.1	189	0.63	7.52	18.0	5.3	1.40	< 0.030	4.6	8.5
18	56.68	172	12	15	7.13	18.9	182	0.90	7.30	27.2	8.5	4.12	< 0.030	7.1	10.5

IMPORT YOUR FILE

Let's import the OCR'ed PDF. Then hit the green "extract data" button.



		1.1.2		cy Oi i	regii					Catil	ent Pl		0050053 03		
		July 2	022		_		Final Eff	luent Resi	ults			Permit# 0	0050853-07	7-00	
Date	Flow to Creek	Total Californ	Fon 3	Enterococci	Dissolved Oxygen (Field Test)	Temp. (Field Test)	Chloride (Reid Test)	Total Phosphorus	pH @15±1°C	Total Suspended Solids	C800	Ammonia Nitrogen	Unionized Ammonia @15e1°C	Total Kijcidahi Nitrogen	Total Nitrogen
	(ML)		(MPN /100mL)		(mg/L)	CO	(mg/L)	(mg P/L)	(5U)	(mg/L)	(mg/L)	(mg N/L)	(mg N/L)	(mg:N/L)	(mg N/L)
01	56.51	178	18		7.35	17.7		0.31		6.0		0.60		2.4	9.3
02	56.53				7.30	17.6		0.36		8.2		1.20		3.2	9.2
03	58.81				7.37	17.8	190	0.35	7.49	8.6	3.4	0.88	< 0.030	2.7	9.9
04	64.02	461	34		6.52	17.6	196	0.54	7.48	12.0		1.20	< 0.030	3.5	10.7
05	68.67	866	47	78	7.15	17.6	185	0.89	7.42	22.6	6.8	1.96	< 0.030	5.1	11.6
06	60.69	1,840	260		7.75	17.8	183	0.92	7.50	23.2	6.5	1.94	< 0.030	5.3	12.4
07	52.52	1,300	98		8.28	17.8	180	0.98	7.48	24.6	6.9	1.06	< 0.030	4.5	11.5
-80	61.96	1,630	161		8.94	18.2		0.69		15.4		0.93		4.0	11.6
09	58.16				7.81	18.5		0.58		12.9		0.52		2,8	10.6
10	58.74				8.17	18.3	180	0.50	7.51	10.8	4.3	0.73	< 0.030	3.0	11.6
11	60.83	461	51	33	8.07	18.1	184	0.60	7.49	13.4	5.0	0.40	< 0.030	3.1	9.1
12	59.75	727	69		7.43	(200)	1000	200	7.48	7.4	223	1.02		3.2	(20)
13	59.75	821	161		7.27	18333	195	1937.7	7.54	26.8	2.77	1.08	< 0.030	4.6	- 20
14	54.90	1,540	56		5.22	10000	110000	77/27/2	7.44	16.4	7.00	2.50	< 0.030	5.5	1000
15	61.28	3,970	118		6.12			0.61		9.4		1.32		4.1	9.3
16	58.48				7.94		670	0.47	200	13.4	1 20	0.80		3.4	200
17	59.70	5000	100		8.18	1,000	0.5372	(2)1772	7.52	18.0	0.000	1.40		4.6	- 100
18	56.68	172	12	15	10000	100000	102007	7200	7.30	27.2	1.000	4.12		7.1	10000
19	133.45	6,870	921		7.28	57757	6,000	7000	7.39	10.2	2,523	0.90		3.2	17773
20	90.21	102	10		9.31	2011	10.7777	27.55	7.53	9.2	500	0.47	< 0.030	2.6	(37)
21	81.68	194	4		8.87	1		10000	7.49	6.2		2.07	< 0.030	4.3	77777
22	68.61	613	65		8.55	01553		0.13		2.8		2.43		4.1	8.6
23	64.36				8.72	100000		0.11		2.7		1.81		3.3	17.00
24	64,77				8.60		1,000	17.000	7.43	2.6		2.03		3.8	100
25	65.36	488	16	1755	7.62	87411	2000	(E00)	7.51	4.0		1.81	< 0.030	3.6	957
26	70.92	649	43	23	4,000	127727	107,00	(2000)	7.52	3.4	2.3	1.71		3.4	1000
27	80.53	109	4		7.27	0.000		30000	7.45	3.2	1111	1.92	< 0.030	3.5	1559
28	75.96	47	2		7.31	275,475	(253)	10000	7.57	2.9	e 20	1.58	< 0.030	3.4	200
29	73.52	121	5		6.69	0.000		0.13		3.4		0.83		2.6	1939
30	72.37				6.84 8.13	1907-00	10000	0.13	7.40	3.3 4.0	10000	1.19	- 0.030	3.0	0.000
ium	70.69 2080.41				8.13	19.7	197	0.12	7.49	4.0	2.1	1.41	< 0.030	3.3	7.6
	2000000				7.65	18.4	182	0.43	7.48	**	4	4	< 0.030	2.7	9
Avg	67.11 52.52				5.22	-		-	7.48	11 2.6		0.40	-	3.7	



@ Copy to Clipboard

Preview of Extracted Tabular Data

56.51	178	18		7.35	17.7		0.31		6.0		0.60			2.4	9.3
56.53				7.30	17.6		0.36		8.2		1.20			3.2	9.2
58.81				7.37	17.8	190	0.35	7.49	8.6	3.4	0.88]	<	0.030	2.7	9.9
64.02	461	34		6.52	17.6	196	0.54	7.48	12.0	3.9	1.20]	<	0.030	3.5	10.7
68.67	866	47	78	7.15	17.6	185	0.89	7.42	22.6	6.8	1.96]	<	0.030	5.1	11.6
60.69	1,840	260		7.75	17.8	183	0.92	7.50	23.2	6.5	1.94]	<	0.030	5.3	12.4
52.52	1,300	98		8.28	17.8	180	0.98	7.48	24.6	6.9	1.06]	<	0.030	4.5	11.5
61.96	1,630	161		8.94	18.2		0.69		15.4		0.93			4.0	11.6
58.16				7.81	18.5		0.58		12.9		0.52			2.8	10.6
58.74				8.17	18.3	180	0.50	7.51	10.8	4.3	0.73]	<	0.030	3.0	11.6
60.83	461	51	33	8.07	18.1	184	0.60	7.49	13.4	5.0	0.40]	<	0.030	3.1	9.1
59.75	727	69		7.43	18.4	188	0.38	7.48	74	4.1	1.02]	<	0.030	3.2	8.4
59.75	821	161		7.27	18.7	195	0.91	7.54	26.8	6.1	1.08]	<	0.030	4.6	9.1
5400	2 - 10	50		5.00	40.7	407	0.00				0.501		0.000		40.0

TADA!*

*YOU'LL
NEED TO
CLEAN UP
THE
FORMATTING
A BIT,
THOUGH

Α	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р
56.51	178	18		7.35	17.7		0.31		6			0.6			2.4
56.53	110	- 10		7.3	17.6		0.36		8.2			1.2			3.2
58.81				7.37	17.8	190	0.35	7.49	8.6		3.4	0.88]	<	0.03	2.7
64.02	461	34		6.52	17.6	196	0.54	7.48	12			1.20]	<	0.03	3.5
68.67	866	47	78	7.15	17.6	185	0.89	7.42	22.6			1.96]	<	0.03	5.1
60.69	1,840	260		7.75	17.8	183	0.92	7.5	23.2			1.94]	<	0.03	5.3
52.52	1,300	98		8.28	17.8	180	0.98	7.48	24.6			1.06]	<	0.03	4.5
61.96	1,630	161		8.94	18.2		0.69		15.4			0.93	1		
58.16				7.81	18.5		0.58		12.9			0.52	1		2.8
58.74				8.17	18.3	180	0.5	7.51	10.8		4.3	0.73]	<	0.03	3
60.83	461	51	33	8.07	18.1	184	0.6	7.49	13.4		5	0.40]	<	0.03	3.1
59.75	727	69		7.43	18.4	188	0.38	7.48	74		4.1	1.02]	<	0.03	3.2
59.75	821	161		7.27	18.7	195	0.91	7.54	26.8		6.1	1.08]	<	0.03	4.6
54.9	1,540	56		5.22	18.7	197	0.62	744	16.4		5	2.50}	<	0.03	5.5
61.28	3,970	118		6.12	18.9		0.61		9.4			1.32	1		4.1
58.48				7.94	19		0.47		13.4			0.8	3		3.4
59.7				8.18	19.1	189	0.63	7.52	18		5.3	1.40}	<	0.03	4.6
56.68	172	12	15	7.13	18.9	182	0.9	7.3	27.2		8.5	4.12)	<	0.03	71
133.45	6,870	921	-	7.28	17.6	102	0.38	7.39	10.2		4.9	0.90}	<	0.03	3.2
90.21	102	10		9.31	17	129	0.34	7.53	9.2		3.6	0.47)	<	0.03	2.6
81.68	194	4		8.87	17.8	175	0.24	7.49	6.2		2.6	2.07]	<	0.03	4.3
68.61	613	65		8.55	18.1		0.13		2.8			2.43	l i		4.1
64.36				8.72	18.2	100	0.11		2.7			1.83			3.3
64.77				8.6	18.6	195	0.12	7.43	2.6]	<	2	2.03}	<	0.03	3.8
65.36	488	16		7.62	18.6	200	0.13	7.51	4		24	1.81]	<	0.03	3.6
70.92	649	43	23	7.88	18.9	196	0.13	7.52	3.4		2.3	1.71]	<	0.03	3.4
80.53	109	4		7.27	19	195	0.11	7.45	3.2]	<	2	1.92]	<	0.03	3.5
75.96	47	2		7.31	19.2	194	0.12	7.57	2.9]	<	2	1.58]	<	0.03	3.4
73.52	121	5		6.69	19.3		0.13		3.4			0.83	3		2.6
72.37				6.84	19.5		0.13		3.3			1.19)		3
70.69				8.13	19.7	197	0.12	7.49	4		2.1	1.41]	<	0.03	3.3

FOR INSTANCE ...

=CONCATENATE(N3,O3)

N	0	P
;	0.03	<0.03
:	0.03	<0.03
<	0.03	<0.03
<	0.03	<0.03
<	0.03	<0.03
:	0.03	<0.03
	0.03	< 0.03
<	0.03	<0.03
	0.03	<0.03
<	0.03	< 0.03

EVEN HARDER CASE SCENARIO

Non-tabular data inside an image-based PDF

Mississippi campaign finance report for "Bill Luckett for Governor Committee"

DOWNLOAD: https://shorturl.at/kEWY4

→ Original source: Mississippi Secretary of State:

Online Campaign Finance Public Search Portal https://cfportal.sos.ms.gov

Reporting Period 8/16/2011 through 9/	30/2011
ITEMIZED DISBU	RSEMENTS
Full Name ActBluc Mississippi	Date Amount of each disbursement this period
Meiling Address 14 Arrow St Ste 11	08/25/2011 \$41.
City, State, Zip Code Cambridge, MA 02138-5106	
Purpose of Dishursement (Optional) Fundraising fees	Aggregate \$1,952.
Full Name ActBlue Mississippi	Date Amount of each (Mo., Day, Year) disbursement this period
Mailing Address 14 Arrow St Stc 11	08/31/2011 \$274.
City, State, Zip Code Cambridge, MA 02138-5106	
Purpose of Disbursement (Optional) Fundraising fees	Aggregate \$1,952.
Full Name AT & T Mobility	Date Amount of each (Mo., Day, Year) disbursement this period
Malling Address PO Box 538695	08/22/2011 \$82.
City, State, Zip Code Atlanta, GA 30353-8695	
Purpose of Disbursement (Optional) cellular phone service	Aggregate \$1,675.
Full Name AT & T Mobility	Date Amount of each (Mo., Day, Year) disbursement this peri
Mailing Address PO Box 538695	08/26/2011 \$137.

THIS CALLS FOR GOOGLE PINPOINT + BETA DATA EXTRACTION TOOL



Sign into Google Pinpoint here:

journaliststudio.google.com/pinpoint/about

Wait for approval (usually very quick)

Request access to to the beta data extraction tool here: shorturl.at/xABU4

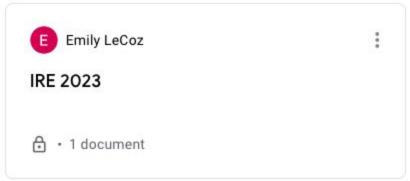
Wait for approval (also relatively quick)

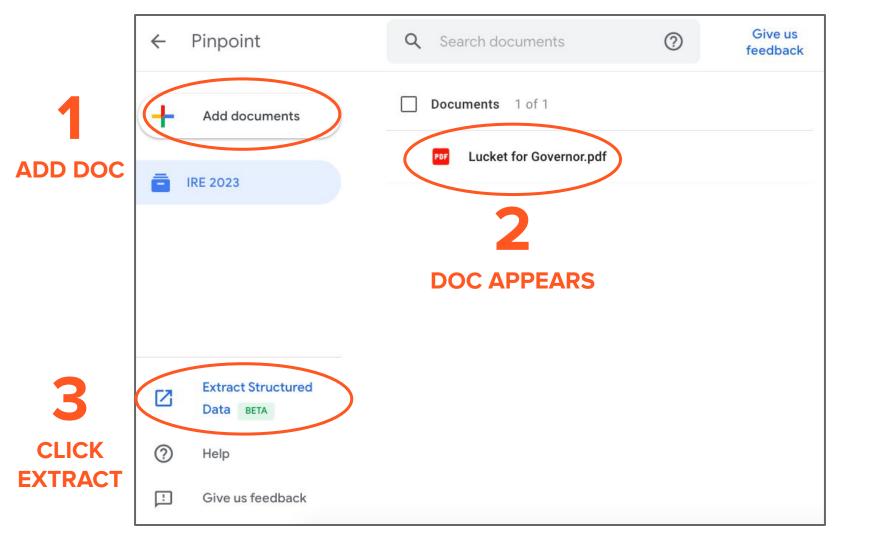
My workspace

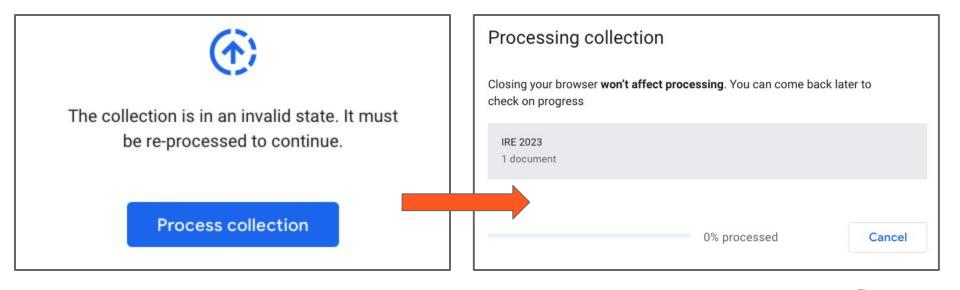
Collections I created or shared with me

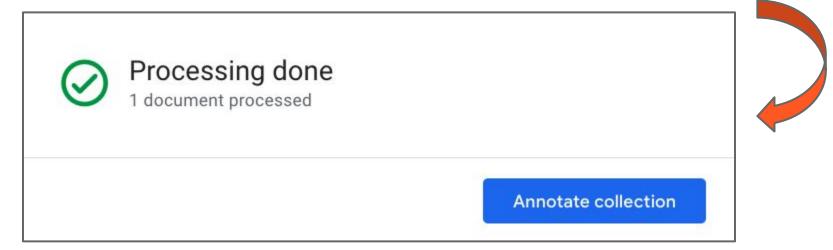
















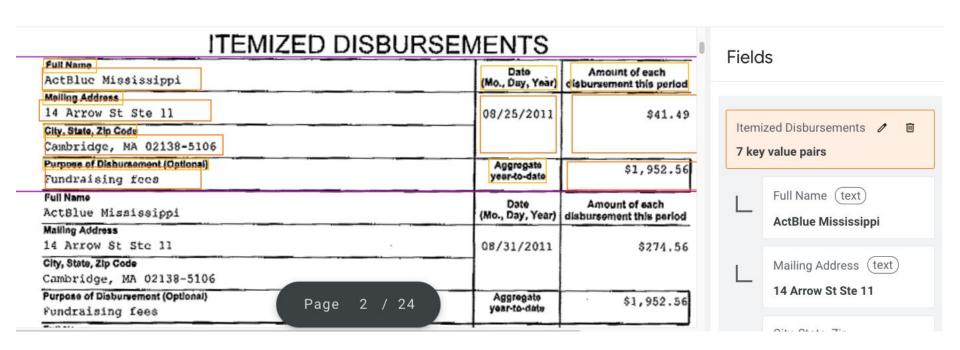








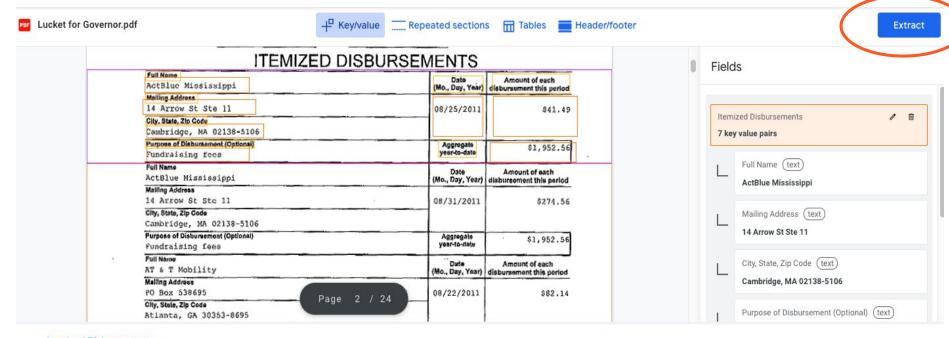




THERE'S A PREVIEW WINDOW UNDER THIS VIEW ...

CHECK THE PREVIEW WINDOW

Full Name	Mailing Address	City, State, Zip Code	Purpose of Disbursement (Optional)	Dato (Mo., Day, Year	
ActBlue Mississippi	14 Arrow St Ste 11	Cambridge, MA 02138-5106	Fundraising fees	08/25/2011	
ActBlue Mississippi	14 Arrow St Ste 11	Cambridge, MA 02138-5106	Fundraising fees	08/31/2011	
AT&T Mobility	PO Box 538695	Atlanta, GA 30353-8695	cellular phone service	08/22/2011	
AT&T Mobility	PO Box 538695	Atlanta, GA 30353-8695	cellular phone service	08/26/2011	
AT&T	PO Box 105262	Atlanta, GA 30348-5262	phone service	08/17/2011	
Bank of America	PO Box 851001	Dallas, TX 75285-1001	airplane fuel	08/23/2011	
DeMixtric Biggs	101 Grandview Dr	Florence, MS 39073	contract services	08/24/2011	
Brent Caldwell	PO Drawer 1000	Clarksdale, MS 38614	contract services,	08/24/2011	
Citi Cards	PO Box 183051	Columbus, OH 43218-3051	tigereye promotions	08/16/2011	



Itemized Disbursements

Document	Full Name	Mailing Address	City, State, Zip Code	Purpose of Disbursement (Optional)	Dato (Mo., Day, Year)	Amount of each disbursement this perio
Lucket for Governor.pdf	ActBlue Mississippi	14 Arrow St Ste 11	Cambridge, MA 02138-5106	Fundraising fees	08/25/2011	\$41.49
	ActBlue Mississippi	14 Arrow St Ste 11	Cambridge, MA 02138-5106	Fundraising fees	08/31/2011	\$274.56

TADA!

File Name	Full Name	Mailing Address	City, State, Zip Code	Purpose of Disbursement (Op D	oato (Mo., Day, Year)	Amount of each disburseme	Aggregate year-to-date
Lucket for Governor.pdf	ActBlue Mississippi	14 Arrow St Ste 11	Cambridge, MA 02138-5106	Fundraising fees	8/25/11		\$1,952.56
Lucket for Governor.pdf	ActBlue Mississippi	14 Arrow St Ste 11	Cambridge, MA 02138-5106	Fundraising fees	8/31/11	\$274.56	\$1,952.56
Lucket for Governor.pdf	AT&T Mobility	PO Box 538695	Atlanta, GA 30353-8695	cellular phone service	8/22/11	\$82.14	\$1,675.40
Lucket for Governor.pdf	AT&T Mobility	PO Box 538695	Atlanta, GA 30353-8695	cellular phone service	8/26/11	\$137.26	\$1,675.40
Lucket for Governor.pdf	AT&T	PO Box 105262	Atlanta, GA 30348-5262	phone service	8/17/11	\$269.13	\$2,360.85
Lucket for Governor.pdf	Bank of America	PO Box 851001	Dallas, TX 75285-1001	airplane fuel	8/23/11	\$768.65	\$11,425.48
Lucket for Governor.pdf	DeMixtric Biggs	101 Grandview Dr	Florence, MS 39073	contract services	8/24/11	\$750.00	\$2,500.00
Lucket for Governor.pdf	Brent Caldwell	PO Drawer 1000	Clarksdale, MS 38614	contract services, mileage	8/24/11	\$2,802,14	\$22,311.29
Lucket for Governor.pdf	Citi Cards	PO Box 183051	Columbus, OH 43218-3051	tigereye promotions	8/16/11	\$331.82	\$3,187.69
Lucket for Governor.pdf	Eleison Group, LLC	3711 Albemarle St NW	Washington, DC 20016-1805	contract services	8/24/11	\$3,000.00	\$35,358.03
Lucket for Governor.pdf	Eleison Group, LLC	3711 Albemarle St NW	Washington, DC 20016-1805	reimbursement for expenses	8/25/11	\$875.81	\$35,358.03
Lucket for Governor.pdf	Will Godfrey	216 Linton Ave	Natchez, MS 39120	contract servics, gas receipt	8/24/11	\$1,115.75	\$16,610.65
Lucket for Governor.pdf	Google	1600 Amphitheatre Parkway	Mountain View, CA 94043	advertising	8/17/11	\$5,500.00	\$20,200.00
Lucket for Governor.pdf	Google	1600 Amphitheatre Parkway	Mountain View, CA 94043	advertising	8/24/11	\$1,000.00	\$20,200.00
Lucket for Governor.pdf	Google	1600 Amphitheatre Parkway	Mountain View, CA 94043	advertising	8/25/11	\$1,000.00	\$20,200.00
Lucket for Governor.pdf	Harvey Dallas Printing & Graphics	PO Box 902	Jackson, MS 39205-0902	walk cards	8/17/11	\$738.30	\$3,499.39
Lucket for Governor.pdf	James Hull	PO Box 213	Tupelo, MS 38802-0213	reimbursement for food costs	8/20/11	\$750.00	\$1,750.00
Lucket for Governor.pdf	Jackson Free Press	PO Box 5067	Jackson, MS 39296	advertising	8/19/11	\$475.00	\$950.00
Lucket for Governor.pdf	Luckett Tyner Law Firm, PA	PO Box 1000	Clarksdale, MS 38614-1000	copies	9/2/11	\$110.40	\$3,070.21
Lucket for Governor.pdf	Francine Luckett	333 Westover Drive	Clarksdale, MS 38614	loan repayment and interest i	9/9/11	\$126,669.40	\$288,877.19
Lucket for Governor.pdf	Whitney Luckett	504 Dyson Street	Fairhope, AL 36532	contract services	8/24/11	\$1,250.00	\$8,817.86
Lucket for Governor.pdf	Regina Watts	1396 Stovall Rd	Clarksdale, MS 38614-9790	accounting services for Augus	8/24/11	\$600.00	\$2,400.00
Lucket for Governor.pdf	Regions Bank	211 East Second St	Clarksdale, MS 38614	wiring fee.	8/17/11	\$20.00	\$263.00
Lucket for Governor.pdf	Regions Bank	211 East Second St	Clarksdale, MS 38614	wiring fee	8/24/11	\$20.00	\$263.00
Lucket for Governor.pdf	Regions Bank	211 East Second St	Clarksdale, MS 38614	bank service charge	8/31/11	\$26.50	\$263.00
Lucket for Governor.pdf	SKDKnickerbocker, LLC	1818 N St NW, Suite 450	Washington, DC 20036-2473	advertising	8/17/11	\$40,000.00	\$280,183.43
Lucket for Governor.pdf	Samantha Styers	1363 Memorial Drive	Cleveland, MS 38732	contract services	8/24/11	\$950.00	\$13,129.29
Lucket for Governor.pdf	The Dinner Bell Restaurant	1000 South Harper Road	Corinth, MS 38834	food for event	8/20/11	\$250.00	\$250.00
Lucket for Governor.pdf	US Bank	PO Box 790108	Saint Louis, MO 63179-0408	lodging, advertising	8/23/11	\$1,213.00	\$2,212.64
Lucket for Governor.pdf	Julienno Uhlich	PO Box 69	Clarksdale, MS 36814	reimbursement for event cost	8/22/11	\$500.00	\$38,468.53
Lucket for Governor.pdf	Julienne Uhlich	PO Box 69	Clarksdale, MS 36814	Election Day expenses	8/23/11	\$350.00	\$38,468.53

VICTORY!



QUESTIONS OR COMMENTS?

EMILY LE COZ elecoz@usatoday.com